

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 16, 2002, 12:53:58 ; Search time 2228 Seconds
(without alignments)
10904.183 Million cell updates/sec

Title: US-09-822-849A-1

Perfect score: 1800
Sequence: 1 acagacagacagacgagcgtt.....ataaaaaaagattacatcc 1800

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 674847542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estlba:*
2: em_estlhum:*
3: em_estlin:*
4: em_estlmu:*
5: em_estlov:*
6: em_estlpl:*
7: em_estlro:*
8: em_estlrc:*
9: gb_estl:*
10: gb_estl2:*
11: gb_hrc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	789.6	43.9	808	10	BG681012
2	781	43.4	887	10	Z78306
3	775.6	43.1	819	9	AI760329
4	773.2	43.0	814	9	AI760329
5	738.6	41.0	829	9	AI355567
6	737.4	41.0	773	9	AI355567
7	717.6	39.9	778	10	BG435706
8	680.8	37.8	847	9	AI159972
9	675.6	37.5	728	9	AM072830
10	670.4	37.2	1080	10	BG028963
11	662.6	36.8	991	10	BE873929
12	657.8	36.5	796	10	BE385520
13	655.4	36.4	908	10	BG113228
14	648	36.0	748	10	BE545112
15	642.8	35.7	729	9	BF798210
16	631.8	35.1	838	9	AI159665
17	630	35.0	726	10	BG258067

c	18	621.4	34.5	863	9	AI159667	AI159667
	19	617.2	34.3	749	10	BE885890	BE885890
	20	602.6	33.5	620	10	BG403149	BG403149
	21	593.4	33.0	689	10	BE881346	BE881346
c	22	591.8	32.9	623	9	AM341128	AM341128
	23	579.4	32.2	621	9	AV682561	AV682561
	24	570	31.7	838	10	BE185652	BE185652
	25	565	31.4	579	10	BE656081	BE656081
	26	563.8	31.3	579	10	BG570725	BG570725
c	27	550	30.6	561	9	BE217917	BE217917
	28	547	30.4	555	10	BF061940	BF061940
c	29	545	30.3	555	9	AI627511	AI627511
	30	544.4	30.2	666	9	AI765113	AI765113
c	31	543	30.2	551	9	AA707321	AA707321
c	32	534	29.7	547	9	AM152664	AM152664
c	33	533.8	29.7	583	9	AM392295	AM392295
c	34	532.8	29.6	762	10	BG170260	BG170260
c	35	531.2	29.5	548	10	BF433363	BF433363
c	36	527	29.3	535	9	AM065558	AM065558
c	37	526.2	29.2	539	9	AM471380	AM471380
c	38	524.6	29.1	577	9	AI369713	AI369713
c	39	522.8	29.0	555	9	AM842572	AM842572
c	40	522.4	29.0	705	10	BI655350	BI655350
c	41	522	29.0	560	10	BE672542	BE672542
c	42	522	29.0	566	9	AI694527	AI694527
c	43	522	29.0	625	9	AI152016	AI152016
c	44	508.4	28.2	632	9	AV702950	AV702950
c	45	503.2	28.0	729	10	BF219148	BF219148

ALIGNMENTS

RESULT 1
LOCUS BG681012 808 bp mRNA linear EST 01-MAY-2001
DEFINITION 602628831f1 NCI_CGAP_Skn4 Homo sapiens cDNA clone IMAGE:4753735 5',
mRNA sequence.
ACCESSION BG681012
KEYWORDS BG681012.1 GI:13912409
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 808)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: James Cleaver, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
plate: LLAM10613 row: n column: 08
High quality sequence stop: 807.
Location/Qualifiers
1..808
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4753735"
/clone_1lb="NCI_CGAP_Skn4"
/tissue="type="squamous cell carcinoma"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: skin; Vector: pCMV-SPOrt6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.5kb. Library constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

FEATURES

source

BASE COUNT

280 a 160 c 210 g 158 t

ORIGIN

Query Match 43.9% Score 789.6; DB 10; Length 808;
 Best Local Similarity 99.4%; Pred. No. 1.4e-128;
 Matches 803; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 573 TTGAGAAATGTAACATTAATGCTGACCGAGAGCTAATGGAATCAATGAGTTGAAA 632
 DB 1 TTGAGAAATGTAACATTAATGCTGACCGAGAGCTAATGGAATCAATGAGTTGAAA 60
 QY 633 GTTAGAACCTTCAGAAAATAAGCTAGCAAAAAGATTGGAACAGAAATGGAATTTGAAA 692
 DB 61 GTTAGAACCTTCAGAAAATAAGCTAGCAAAAAGATTGGAACAGAAATGGAATTTGAAA 120
 QY 693 TTGAGCCAGATAAAGATTAATCCCTTCTCTCGGAAAGAAATGTCAGTCTTTAG 752
 DB 121 TTGAGCCAGATAAAGATTAATCCCTTCTCTCGGAAAGAAATGTCAGTCTTTAG 180
 QY 753 ACATGAAAAGAGCTTCAGAGAAAAGAAAAGAAATATGAGCCCAACCTGAGCCTG 812
 DB 181 ACATGAAAAGAGCTTCAGAGAAAAGAAAAGAAATATGAGCCCAACCTGAGCCTG 240
 QY 813 TGGCTCAACCTCAGCCTCAGTCTCAGCCCGAGCTTCAGTCTCAATCCAGTCCACAG 872
 DB 241 TGGCTCAACCTCAGCCTCAGTCTCAGCCCGAGCTTCAGTCTCAATCCAGTCCACAG 300
 QY 873 TACTCAGATCCAGCCTCCTCTCAGCCTGAGAGATTGTCATTAGCTGTTTACAGCAA 932
 DB 301 TACTCAGATCCAGCCTCCTCTCAGCCTGAGAGATTGTCATTAGCTGTTTACAGCAA 360
 QY 933 CACCCCAAGTTACTAGAGAGCAAGGCAATTACTAGAGAGAGAGATTTCCTGTAG 992
 DB 361 CACCCCAAGTTACTAGAGAGCAAGGCAATTACTAGAGAGAGAGATTTCCTGTAG 420
 QY 993 AGTCTGTAACCTCAGGAGTACAGTACAGCCAGTCTGACAGTACATCCAGAGCA 1052
 DB 421 AGTCTGTAACCTCAGGAGTACAGTACAGCCAGTCTGACAGTACATCCAGAGCA 480
 QY 1053 AGAGCAAAACCAAACTAGAGAGCAAGAGTACAGTACAGTACAGTACAGTACAG 1112
 DB 481 AGAGCAAAACCAAACTAGAGAGCAAGAGTACAGTACAGTACAGTACAGTACAG 540
 QY 1113 GTGAAAGTGAAGAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAG 1172
 DB 541 GTGAAAGTGAAGAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAG 600
 QY 1173 GAAGTAGTTCACAGAGTGAAGTAGTAGTACAGTACAGTACAGTACAGTACAGTAC 1232
 DB 601 GAAGTAGTTCACAGAGTGAAGTAGTAGTACAGTACAGTACAGTACAGTACAGTAC 660
 QY 1233 CAAGTGGACACAGCAGAGAGATAGTAGCAGTAGCAGTACAGTACAGTACAGTACAG 1292
 DB 661 CAAGTGGACACAGCAGAGAGATAGTAGCAGTAGCAGTACAGTACAGTACAGTACAG 720
 QY 1293 GTGAGAGTAGGGGTGCGGGACATTAATAGAGATGAAGAGCAGAGAGGCTGATCGG - 1351
 DB 721 GTGAGAGTAGGGGTGCGGGACATTAATAGAGATGAAGAGCAGAGAGGCTGATCGGA 780
 QY 1352 AAGAGAAAGGATACTTCAGAGTACAGAA 1379
 DB 781 AAGAGAAAGGATACTTCAGAGTACAGAA 808

RESULT 2
 278306 887 bp mRNA linear EST 28-JUL-1999
 LOCUS HS278306 Human fetal brain S. Meier-Ewert Homo sapiens cDNA clone
 DEFINITION 1.38 (CEPH) 3', mRNA sequence.
 ACCESSION 278306
 VERSION 278306.1 GI:1495079
 KEYWORDS EST.
 SOURCE human,
 ORGANISM Homo sapiens

REFERENCE
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 Neri, C., Albanese, V., Lebre, A.S., Holbert, S., Saada, C., Bouguetier, L., Meier-Ewert, S., Legall, I., Millaudeau, P., Bul, H., Guindon, J., Massart, C., Guillou, S., Gervy, P., Poullier, E., Rigault, P., Weissenbach, J., Lennon, G., Chumakov, I., Dausset, J., Lehrach, H., Cohen, D. and Cann, H.M.
 Survey of CAG/CTG repeats in human cDNAs representing new genes: candidates for inherited neurological disorders
 Hum. Mol. Genet. 5 (7), 1001-1009 (1996)
 96414310

JOURNAL
 MEDLINE
 COMMENT
 Contact: Neri C.
 Fondation Jean Dausset - CEPH
 27 Rue Juliette Dodu, 75010 Paris, France
 Related sequence: 278305
 3' - sequence (lower strand).
 Location/Qualifiers
 1..887
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="Human fetal brain S. Meier-Ewert"
 /tissue_type="brain"
 /dev_stage="fetus"
 /note="cDNA library of S. Meier-Ewert, Max Planck Inst. f. Mol. Genetics, Berlin, FRG"

BASE COUNT 198 a 222 c 167 g 288 t 12 others
 ORIGIN

Query Match 43.4% Score 781; DB 10; Length 887;
 Best Local Similarity 96.3%; Pred. No. 4.6e-127;
 Matches 856; Conservative 0; Mismatches 27; Indels 6; Gaps 6;

QY 868 ACCAGTACTCCAGTCCAGCTCCT-CTCAGCCTGAGATTGTCATTAGCTGTTTAC 926
 DB 886 ACCAGTACTCCAGTCCAGCTCCT-CTCAGCCTGAGATTGTCATTAGCTGTTTAC 827
 QY 927 AGCCAAACCCCAAGTACTCAGAGCAAGGCAATTACTACTGAGAGAGAAATTTTC 986
 DB 826 AGCCAAACCCCAAGTACTCAGAGCAAGGCAATTACTACTGAGAGAGAAATTTTC 767
 QY 987 CTGTAGAGTCTGTAACCTCAGTACAGTACAGTACAGTACAGTACAGTACAGTACAG 1046
 DB 766 CTGTAGAGTCTGTAACCTCAGTACAGTACAGTACAGTACAGTACAGTACAGTACAG 707
 QY 1047 AAGCAAGCAAAACCAAACTAGAGAGCAAGAGTACAGTACAGTACAGTACAGTAC 1106
 DB 706 AAGCAAGCAAAACCAAACTAGAGAGCAAGAGTACAGTACAGTACAGTACAGTACAG 647
 QY 1107 GCAGAGTAGAAGTCAAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAG 1166
 DB 646 GCAGAGTAGAAGTCAAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAG 587
 QY 1167 GCAGTGAAGTATTCACAGAGTGAAGTAGTAGCAGTGCAGTAGTCCAGTAGCAGC 1225
 DB 586 GCAGTGAAGTATTCACAGAGTGAAGTAGTAGCAGTGCAGTAGTCCAGTAGCAGC 527
 QY 1226 TCCAGTACAAAGTGGGCGACAGCAGCAGAGTATAGCAGTACAGTACAGTACAGTACAG 1285
 DB 526 TCCAGTACAAAGTGGGCGACAGCAGCAGAGTATAGCAGTACAGTACAGTACAGTACAG 467
 QY 1286 AGTAGAAGTGGAGTAGAGGGGTGCGGACATTAATAGAGATGAAGAGCAGAGAGGCTG 1345
 DB 466 AGTAGAAGTGGAGTAGAGGGGTGCGGACATTAATAGAGATGAAGAGCAGAGAGGCTG 409
 QY 1346 GATGGAAGAGAGGATCTTCAGAGTACAGAAAGTACAAATCTTCAAAAGGTGT 1405
 DB 408 GATGGAAGAGAGGATCTTCAGAGTACAGAAAGTACAAATCTTCAAAAGGTGT 349
 QY 1406 GGAGTAGAGATA-CAAAAGATCAAAAGATTAAGAAATTCGGGTCCGACAGAAAGGTCT 1464
 DB 348 GGAGTAGAGATNCCAAAGATCAAAAGATTAAGAAATTCGGGTCCGACAGAAAGGTCT 290

Helix Research Institute
 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 Tel.: 81-438-52-3951
 Fax: 81-438-52-3952
 Email: genomics@hri.co.jp
 HRI human cDNA project; 5' - & 3' -end one pass sequencing: Helix

RESULT 10	
LOCUS	BC028963
DEFINITION	1080 bp mRNA linear EST 24-JAN-2001
ACCESSION	602293090f1 NIH_MGC_86 Homo sapiens CDNA clone IMAGE:4387558 5',
VERSION	BC028963
KEYWORDS	BC028963.1 GI:12418058
SOURCE	EST.
ORGANISM	human.
	Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 1080)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
plate: LLAM10072 row: d column: 23
High quality sequence stop: 643.

source	BASE COUNT	ORIGIN
1. 1080	378 a	177 c 312 g 213 t
/organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:4387558" /clone_id="NIH_MGC_86" /tissue_type="osteosarcoma, cell line" /lab_host="DH10B (phage-resistant)" /note="Organ: bone; Vector: pCMV-SPORT6; Site.1: NotI; Site.2: SalI; cloned unidirectionally; oligo-dT primed. Average insert size 1.533 kb. Library enriched for full-length clones and constructed by Life Technologies Note: this is a NIH_MGC library."		

Query Match	37.2%;	Score 670.4;	DB 10;	Length 1080
Best Local Similarity	91.9%;	Pred. No. 1.1e-107;		

	Matches	730;	Conservative	0;	Mismatches	61;	Indels	3;	Gaps	2;
0Y	407	GAGGAAAAAGGAAT	TACCGATTGCTTCACTAGTGCAGGAGAAAAACAAGAGAGGAAAGA	466						
Db	1	GAGGAAAAAGGAAT	TACCGATTGCTTCACTAGTGCAGGAGAAAAACAAGAGAGGAAAGA	60						
0Y	467	CAAAAAACGAAAA	TGAGGTTAAGATGAGGAGGAGAACTGAGTTAAGGGAAAGTGGAAG	526						
Db	61	CAAAAAACGAAAA	TGAGGTTAAGATGAGGAGGAGAACTGAGTTAAGGGAAAGTGGAAG	120						
0Y	527	CACGAGATAGTAC	GCGCTTGAGAAAGTTATGAGATGTGCTGAGATGGTTGAGAAATGTCAA	586						
Db	121	CACGAGATAGTAC	GCGCTTGAGAAAGTTATGAGATGTGCTGAGATGGTTGAGAAATGTCAA	180						
0Y	587	CATGTAAATGCTG	ACCAGAGGTAATGGAACATAATCGAGTTGAAAGTTAGAACCTTCA	646						
Db	181	CATGTAAATGCTG	ACCAGAGGTAATGGAACATAATCGAGTTGAAAGTTAGAACCTTCA	240						
0Y	647	GAATAATGAACTG	CAAAATATGGAACAGAAATGAAATTTGAAATTTAGCCAGATATA	706						
Db	241	GAATAATGAACTG	CAAAATATGGAACAGAAATGAAATTTGAAATTTAGCCAGATATA	300						
0Y	707	GAATGTAAATCC	TTCTCTCGTGGAAAGAAATGTACGTAGCTTGTAGACATGAAAAAGAG	766						
Db	301	GAATGTAAATCC	TTCTCTCGTGGAAAGAAATGTACGTAGCTTGTAGACATGAAAAAGAG	360						
0Y	767	TCTGAGGAAAAAG	AAAGAAATCATCTGAGCCCAACTGAGCGCTGGCTCAACCTGAG	826						
Db	361	TCTGAGGAAAAAG	AAAGAAATCATCTGAGCCCAACTGAGCGCTGGCTCAACCTGAG	420						
0Y	827	CCCTCAATCTCAG	CCCCAGCTTCAAGCTTCAATCCCACTCCCAACCACTACTCCAGTCCAG	886						
Db	421	CCCTCAATCTCAG	CCCCAGCTTCAAGCTTCAATCCCACTCCCAACCACTACTCCAGTCCAG	480						
0Y	887	CCCTCCCTCAGC	CTGAGGATTTGTTCATTAAGTGTGTTTACAGCCAAACCCCAAGTTACT	946						
Db	481	CCCTCCCTCAGC	CTGAGGATTTGTTCATTAAGTGTGTTTACAGCCAAACCCCAAGTTACT	540						
0Y	947	CAGGACAAAGGCA	TTTATCTAGAGGAAAGATTTCCCTGTAGAGCTCTATAAATCTC	1006						
Db	541	CAGGACAAAGGCA	TTTATCTAGAGGAAAGATTTCCCTGTAGAGCTCTATAAATCTC	600						
0Y	1007	ACTGAGATCAC	AGTAGGCACTTTGACAGTATCCAGAGAGCAAGGCAAAACCAA	1066						
Db	601	CTGAGGTATCA	CAAGTAGAGCCAGCTTGTGACAGTACATTGACAGGCAAGGAAAGAAC - AA	658						
0Y	1067	ACTAGAGCAGA	AGTAGAGGTGAGACTGTGAATTAACAAGCAACAGTAAGAAATGGAAGC	1126						
Db	659	AATAGAGCAGA	AGTAGAGGTGAGACTGTGAATTAACAAG - GAAGAGTTGAAGTGAAGA	717						
0Y	1127	ACTAGCAGTAG	CAGTTCATAGTAGCAGTTCCMACCCAGTAGCAGCAGTAGAGTAGTATCCAGC	1186						
Db	718	CAGTAGCAGTAT	CAGTTCATAGTAGCAGTTCACACAGTGTGAGCGGTGGGAAATGTCCAGAG	777						
0Y	1187	AGTGAAGTAG	TAG 1200							
Db	778	GGGAGTATTG	CAAG 791							

RESULT	11
BE873929	
LOCUS	
DEFINITION	BE873929 991 bp mRNA linear EST 20-OCT-2000 601484590071 NIH_MGC_69 Homo sapiens cDNA clone IMAGE:386822 5'
ACCESION	mRNA sequence.
VERSION	BE873929
KEYWORDS	BE873929.1 GI:10322705
SOURCE	EST.
ORGANISM	human. Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 991) NIH-MGC http://mgc.ncl.nih.gov/ .

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 16, 2002, 12:50:53 ; Search time 3392 Seconds

(without alignments)
11104.884 Million cell updates/sec

Title: US-09-822-849A-1

Perfect score: 1800
Sequence: 1 acagacagaaactgcgcctt.....ataaaaaaattacatcc 1800

Scoring table: IDENTITY_NUC
Gapop 10.0 ; Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : GenBankl:*

- 1: gb_da:*
- 2: gb_hlg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vl:*
- 15: em_da:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vl:*
- 30: em_hlg_hum:*
- 31: em_hlg_inv:*
- 32: em_hlg_other:*
- 33: em_higo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
------------	-------	-------	-------	--------	----	----	-------------

1	1787.2	99.3	2956	9	HSSDK3	Y10351 H. sapiens m
2	1787.2	99.3	3122	9	AF112222	AF112222 Homo sapi
3	1777.6	98.8	3278	9	HSMEMA	Y09703 H. sapiens m
4	1730.4	96.1	2369	6	AR136786	AR136786 Sequence
5	1646.6	91.5	2617	7	HSU77718	U77718 Human desmo
6	1603.4	89.1	196319	9	CNSOLDT4	AL132639 Human chr
7	1597	88.7	11787	9	AF195139	AF195139 Homo sapi
8	1406.6	78.1	2813	10	MMPININ	Y08701 M. musculus
9	1394	77.4	3893	10	CRU77716	U77716 Canis famli
10	1265	70.3	151836	10	AC09389	AC09389 Rattus no
11	1067	59.3	4167	4	BU077717	U77717 Bos taurus
12	806.8	44.8	1109	4	PIGSERNED	M55701 Pig neutrop
13	470	26.1	62685	2	AC100145	AC100145 Mus muscu
14	468.4	26.0	49360	2	AC100145	AC100145 Rattus no
15	387.2	21.5	2306	5	XLSDK2	G25957 Human STS E
16	385	21.4	2402	5	XLSNUCLEAR	AX197800 Sequence
17	339.4	18.9	349	5	HS059479	AX208402 Sequence
18	320	17.8	62685	2	AC100145	G54093 914 Human H
19	308.2	17.1	181089	2	AC094069	AC090533 Mus muscu
20	197.8	11.0	461	11	G25957	AF13051 Zea mays
21	166.4	9.2	239	6	AX197800	E10125 DNA encodin
22	166.4	9.2	239	6	AX208402	E10125 DNA encodin
23	143	7.9	143	11	G54093	AJ271723 Fugu rubr
24	139.8	7.8	27245	2	AC090533	AL662884 Homo sapi
25	136.4	7.6	666	8	AF13050S2	AL662884 Homo sapi
26	136.4	7.6	2069	23	E10125	AF13050 Zea mays
27	136.4	7.6	3399	23	E10125	AF397467 Ictalurus
28	136.2	7.6	22398	5	FRU271723	AC097839 Rattus no
29	136	7.6	144505	2	AL662884	AL606925 Mus muscu
30	133.2	7.4	282611	2	AL645746	MI6230 Strongyloce
31	132.4	7.4	543	8	AF13050S1	AC020858 Mus muscu
32	132.4	7.4	13661	14	U93872	G20375 Human STS A
33	130.4	7.2	10115	5	AF397467	AY016024 Takifugu
34	130	7.2	98221	2	AC097839	I66494 Sequence 14
35	129.2	7.2	237258	2	AL606925	AC096822 Rattus no
36	127.8	7.1	14091	3	SUSMPL	AF305694 Rattus no
37	127.6	7.1	186222	2	AC020858	AC098456 Rattus no
38	127.4	7.1	230	11	G20375	AF13054 Zea mays
39	127	7.1	35793	5	AY016024	AF139019 Cepaea ne
40	125.6	7.0	7218	6	I66494	
41	125.4	7.0	17033	2	AC096822	
42	123.8	6.9	3127	14	AF305694	
43	123.6	6.9	257302	2	AC098456	
44	123.2	6.8	693	8	AF13054	
45	122.8	6.8	624	3	AF139019	

ALIGNMENTS

RESULT 1
HSSDK3 LOCUS 2956 bp mRNA linear PRI 02-APR-1998
DEFINITION H.sapiens mRNA for nuclear protein SDK3, partial.
ACCESSION Y10351
VERSION Y10351.1 GI:3021391
KEYWORDS nuclear protein SDK3.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Brandner,J.M., Reidenbach,S. and Franke,W.W.
TITLE Evidence that 'p1in', reportedly a differentiation-specific desmosomal protein, is actually a widespread nuclear protein
JOURNAL JOURNAL MEDLINE
MEDLINE 96109012
REFERENCE 2 (bases 1 to 2956)
AUTHORS Brandner,J.M.
TITLE Direct Submission
JOURNAL Submitted (06-JAN-1997) J.M. Brandner, German Cancer Research Centre, Cell Biology, Im Neuenheimer Feld 280, 69120 Heidelberg, FRG

FEATURES
source

Location/Qualifiers
1..2956
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="H22"
/cell_line="Hacat"
/dev_stage="adult"
<1..2138
/codon_start=3
/product="nuclear protein SDK3"
/protein_id="CAN71377.1"
/db_xref="GI:3021392"
/db_xref="SPTREMBL:O60899"
/translation="TLEQLLEKAKESIKLVNENIRKLTGRDPNDVPIGARLTALSGP
GGGRGSLILRRGFSDSGGPPAKODLEGAVSRIGERRTRESRSEDDEDDVK
KPLQSSVATSKERTRDILPONNDEKGNRRIFGLMTLOKFOESTVATER
OKRROILOKLEQAEERKOVNERBELFEERAROTELRIEOKVELLAQOENNE
HNAKIIKYLRTKRLPLFYIRGMBPATORLIESORKNALFGRIRIAPADINME
ARRRSMKREKHOVYRNREOKAEOEBCKVAOEBELEETGNHNVLEAGVEEEK
ELAVISDEKBOEEEDKOEMEYMEETEVESERKQDSOEYVADVLEENVEKH
VIADQVMEETNRESEVSENEASKLEPEMEFEIPEDKSLSGKENVALDMEK
ESEKEKESEPEPEVADPOPOPOLOQSOQVLOQSPQSEDLSLAVLQTP
QYTOEGHLLPERKDFPVESEVLTLEVPERVLTVHSESKRTIRSRSGRAKNTSK
SRSS
SRSRSGRGNRRKRRSYDRKRDCTGLERHSKSSKGSSSRDTRGSDKNSDRK
RSISPSRSKRSRSEDRKSRDKRR"

CDS

BASE COUNT 979 a 570 c 789 g 618 t
ORIGIN

Query Match 99.3%; Score 1787.2; DB 9; Length 2956;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1792; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 ACACACGAACTGGCGCTTTTGGACAGAAAGTTGACCTCCGACGTGCAAGAAATG 60
DB 590 ACACACGAACTGGCGCTTTTGGACAGAAAGTTGACCTCCGACGTGCAAGAAATG 649
QY 61 GAATGACATATATCCAAATTAATATATATAGAACTAGCAAAAGCCCATTTGTT 120
DB 650 GAATGACATATATCCAAATTAATATATATAGAACTAGCAAAAGCCCATTTGTT 709
QY 121 TTATATTCCTGGAAGATGTCTCAGCTACCCAAAATAATAGAGAGTACAGACAAA 180
DB 710 TTATATTCCTGGAAGATGTCTCAGCTACCCAAAATAATAGAGAGTACAGACAAA 769
QY 181 AATGACGCTTATTTGAAGTAGACCATGATTTGCAAGCAATTAATAATATGA 240
DB 770 AATGACGCTTATTTGAAGTAGACCATGATTTGCAAGCAATTAATAATATGA 829
QY 241 GGCTAGGCTAGAAACATCAATGAAGAAAGAGCATCAGTGGTGGCTTAATGAAGA 300
DB 830 GGCTAGGCTAGAAACATCAATGAAGAAAGAGCATCAGTGGTGGCTTAATGAAGA 889
QY 301 ACACAGGCGGACACAGAGAGGTTAAGTGGCTCAGCGAGAGAAAGAGTTGGAGAGAC 360
DB 890 ACACAGGCGGACACAGAGAGGTTAAGTGGCTCAGCGAGAGAAAGAGTTGGAGAGAC 949
QY 361 AGGTAATCAGCAATGATGTAGAAATAGAGAAAGAGAGAGAGAGAGAAAGAAAT 420
DB 950 AGGTAATCAGCAATGATGTAGAAATAGAGAAAGAGAGAGAGAGAGAGAAAGAAAT 1009
QY 421 AGCATTTCTCATAGTATGAGAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAG 480
DB 1010 AGCATTTCTCATAGTATGAGAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAG 1069
QY 481 GGAGGTTAAGTGAAGAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 540
DB 1070 GGAGGTTAAGTGAAGAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1129
QY 541 GCCTGAAGAGTATGATGTCTAGAGATGTTGAGAAATGTCAACATGTATTTCTGA 600
DB 1130 GCCTGAAGAGTATGATGTCTAGAGATGTTGAGAAATGTCAACATGTATTTCTGA 1189

QY 601 CCAGAGGTAATGAAACTAATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 660
DB 1190 CCAGAGGTAATGAAACTAATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1249
QY 661 CAAGAATTTGAAACAGAAATGAAATTTGAAATTTGAAATTTGAAATTTGAAATTTGAAAT 720
DB 1250 CAAGAATTTGAAACAGAAATGAAATTTGAAATTTGAAATTTGAAATTTGAAATTTGAAAT 1309
QY 721 TTCTCTGGGAAAGAAATGCTGCTTATGACATGAAAGAGAGTGTGAGGAAAGAAAG 780
DB 1310 TTCTCTGGGAAAGAAATGCTGCTTATGACATGAAAGAGAGTGTGAGGAAAGAAAG 1369
QY 781 AGAAAAAGATCTGAGACCCCAACCTGAGCTGTGCTCAACCTGAGTCTGAGCC 840
DB 1370 AGAAAAAGATCTGAGACCCCAACCTGAGCTGTGCTCAACCTGAGTCTGAGCC 1429
QY 841 CCAGCTCAGCTTCAATCCAGTCCCAACAGATGACAGTCCAGCTCCAGCTCCAGCTCCAGCT 900
DB 1430 CCAGCTCAGCTTCAATCCAGTCCCAACAGATGACAGTCCAGCTCCAGCTCCAGCTCCAGCT 1489
QY 901 TGAGGATTTGTCAATTAGCTTTTACAGCCCAACCCCAAGTTACTCAGAGCAAGGCA 960
DB 1490 TGAGGATTTGTCAATTAGCTTTTACAGCCCAACCCCAAGTTACTCAGAGCAAGGCA 1549
QY 961 TTACTACCTGAGAGAGAGATTTCTGTAGAGTCTGTAAACTGACAGTACCAAGT 1020
DB 1550 TTACTACCTGAGAGAGAGATTTCTGTAGAGTCTGTAAACTGACAGTACCAAGT 1609
QY 1021 AGAGCCAGCTTGTGACATGATATCCAGAGACCAAGAGCAAAACCAAACTAGAGAGCAAG 1080
DB 1610 AGAGCCAGCTTGTGACATGATATCCAGAGACCAAGAGCAAAACCAAACTAGAGAGCAAG 1669
QY 1081 TAGAGTTCAGCTGAAATTAATTAACAGCAAGATGAAAGTGAAGAGTGAAGTGAAG 1140
DB 1670 TAGAGTTCAGCTGAAATTAATTAACAGCAAGATGAAAGTGAAGAGTGAAGTGAAG 1729
QY 1141 TTCTAGTACAGTTCACCAAGTACAGCAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1200
DB 1730 TTCTAGTACAGTTCACCAAGTACAGCAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1789
QY 1201 CAGTCGAGTATGTTCCAGTACAGCTTCAATGAGTGGAGAGAGAGAGAGAGATGATG 1260
DB 1790 CAGTCGAGTATGTTCCAGTACAGCTTCAATGAGTGGAGAGAGAGAGAGAGATGATG 1849
QY 1261 CAGTACAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1320
DB 1850 CAGTACAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1909
QY 1321 AGATAGAAAGCAGAGAGAGCGGTGATCGGAAGAGAGAGAGAGAGAGAGAGAGAG 1380
DB 1910 AGATAGAAAGCAGAGAGAGCGGTGATCGGAAGAGAGAGAGAGAGAGAGAGAGAG 1969
QY 1381 AAGTCACAAATCTCAAAAGGTGTGTAGTAGATGATGATGATGATGATGATGATGATG 1440
DB 1970 AAGTCACAAATCTCAAAAGGTGTGTAGTAGATGATGATGATGATGATGATGATGATG 2029
QY 1441 TTCCGGTCCGACAGAAAGAGTCTATATCAGAGAGTATGATCAGGCAAAAGATCTTC 1500
DB 2030 TTCCGGTCCGACAGAAAGAGTCTATATCAGAGAGTATGATCAGGCAAAAGATCTTC 2089
QY 1501 AAGAGTGAAGAGCCGCAAAATCAGACAGAAAGCAAAAGCGTTAATGAGAGAGAGCC 1560
DB 2090 AAGAGTGAAGAGCCGCAAAATCAGACAGAAAGCAAAAGCGTTAATGAGAGAGAGCC 2149
QY 1561 AGGCTTCTTAGCCATCTTTGACAGCAAGATTTCTGTATTAAGAAAGATTTCTCTTCC 1620
DB 2150 AGGCTTCTTAGCCATCTTTGACAGCAAGATTTCTGTATTAAGAAAGATTTCTCTTCC 2209
QY 1621 TTGTAAAGAGATGCTGCTTATGAGATGATGATGATGATGATGATGATGATGATG 1680
DB 2210 TTGTAAAGAGATGCTGCTTATGAGATGATGATGATGATGATGATGATGATGATG 2269
QY 1681 GACTGTTGTTTACACAGACATTTCTGACTTTTGCATATTTTGTAAAGATTAATTATC 1740

Db	2270	GACGTGTTGTACCAGACATCTTCTGACTTTTTCATAAATTTTGAAGAGTTATTATTC	2329
1741	AAATATGTAGGTTCCAAATATGTAAAAATGATAATTAATAAAAAAGATTAACATCC	1800	
2330	AAATATGTAGGTTCCAAATATGTAAAAATGATAATTAATAAAAAAGATTAACATCC	2389	

RESULT 2	AF112222	AF112222	3122 bp	RNA	linear	PRI 13-DEC-1999
LOCUS	AF112222					
DEFINITION	Homo sapiens nuclear protein SDK3 mRNA, complete cds.					
ACCESSION	AF112222					
VERSION	AF112222.1	GI:6563229				
KEYWORDS						
SOURCE	human.					
ORGANISM	Homo sapiens					

PERIPHERAL AUTHORS	TITLE	REFERENCE AUTHORS	TITLE	JOURNAL
1 (bases 1 to 3122) Gu, J., Shi, J., Ren, S., Fu, G., Wang, Y., Chen, Z. A novel gene expressed in human adrenal gland	Jin, W., Fu, S., Huang, Q., Dong, H., Yu, Y., Fu, G., Wang, Y., Chen, Z. Unpublished 2 (bases 1 to 3122) Gu, J., Shi, J., Ren, S., Fu, G., Wang, Y., Chen, Z. Direct Submission	Jin, W., Fu, S., Huang, Q., Dong, H., Yu, Y., Fu, G., Wang, Y., Chen, Z. Unpublished 2 (bases 1 to 3122) Gu, J., Shi, J., Ren, S., Fu, G., Wang, Y., Chen, Z. Direct Submission	Chinese National Human Genome Center at Shanghai, Gu Shoujing Rd. 351, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China Location/Qualifiers 1. 3122	

```

      81..2234
      /codon_start=1
      /evidence=not experimental
      /product="nuclear protein SKR3"
      /protein_id="AAFI7209.1"
      /db_xref="GI:563130"
      /translation="MAVAAVRTIQEQLERAKESLKNVDENIKRLTGPDNDVPRIOARI
      LALSGPGGGRGSLILRRGSDSGPPAPRQDLEGVAVSLGGRTRRRSROESDID
      EDDVGRPALOSSVYATSKERTRRDLIDQNMDEKQNRRIFFGLMGTLOKFOEE
      TVALTEKROKREIOETKLEVOAEKEERQVQENRRLFEERRAKOTELRLKRVELAO
      QEAVENENAKIKIKITRTKPLHLYIPRMRPAQOKLEESQRKNALFEERRILEAFI
      OIINMEAPRQSMKEKHQVYRNBEQABQDEGVAKREELBETGQHMVLEIEE
      GEERKEELATVHSDAEKEQEEBEQKQENVEAMEETEVRESEKQDDSPQEEVMDVLE
      VGNKHHYIADQVAKENVESEVPSENASKELEBEMFETLIPDECKSLSGKPNVY
      ALDMKEESEKEKESESEPOPEPPAPQPOSLQPOLQVQPSLOQSPQEDLSL
      VLOLTPQVTOEGHLLPRKDPQVSELTVEPVEPVLYVHPESKSKTRTSRSGRL
      RNKTSKSRSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS
      TSSSESRSRSGRGRGHNRDRKHRRSVYDKRRDTSLELSEHKSXSKGSSRDITKGSXDKKK
      SRSDDKRSISSESSSGKRSRSSESRDKDRDKRR

```

	Query Match	Similarity	99.3%	Score	187.2	DB	9	Length	3122
	Best Local	Similarity	99.6%	Pred.	No.	0			
	Matches	1792	Conservative	0	Mismatches	8	Indels	0	Gaps
QY	1	ACAGACAGAACTGCGGCTTTTGGAAACAGAAAGTTGAGCTTGCGCAGCTCCAAAGAAATG	60						
DB	686	ACAGACAGAACTGCGGCTTTTGGAAACAGAAAGTTGAGCTTGCGCAGCTCCAAAGAAATG	745						
QY	61	GAATGACATAATGCCAAATTAATTAATATATAGAACTAAGACAAAGCCCCATTGTT	120						
DB	746	GAATGACATAATGCCAAATTAATTAATATATAGAACTAAGACAAAGCCCCATTGTT	805						
QY	121	TTATATTCCTGGAAAGATGTGTCAGCTACCCAAAACATATATGAAGTACACAGAAA	180						
DB	806	TTATATTCCTGGAAAGATGTGTCAGCTACCCAAAACATATATGAAGTACACAGAAA	865						

OY	181	AATTAACGCTTTTATTTGAAGGTGACGCATTCGAATTTTGAGAAACAAATTAATTAATTAATGA	240
Db	866	AATGAACGCTTTTATTTGAAGGTGACGCATTCGAATTTTGAGAAACAAATTAATTAATTAATGA	925
OY	241	GGCTAGCGCTTAGAAGACAAATCAATGAAGAAAAAGACATCAGTGTGCTGAATGAAGA	300
Db	926	GGCTAGCGCTTAGAAGACCAATCAATGAAGAAAAAGACATCAGTGTGCTGAATGAAGA	985
OY	301	ACGAACGGCGGAACCAAGAAAGGGGTAAAGGTGGCTCAGCAGAGGGAAGGTTGGAGGAGAC	360
Db	986	ACGAACGGCGGAACCAAGAAAGGGGTAAAGGTGGCTCAGCAGAGGGAAGGTTGGAGGAGAC	1045
OY	361	AGGTAATCAGCACACAAATGATGTAGAAATTAGAGAAAGCAGAGAGAGGAAGGAAAAAGAAAT	420
Db	1046	AGGTAATCAGCACACAAATGATGTAGAAATTAGAGAAAGCAGAGAGAGGAAGAAAAAGAAAT	1105
OY	421	AGCGATTGTTTCATATGATGTATGCAGAGAAAGACAGAGAGGAAGAACAAAAACAGAAAT	480
Db	1106	AGCGATTGTTTCATATGATGTATGCAGAGAAAGACAGAGAGGAAGAACAAAAACAGAAAT	1165
OY	481	GGAGGTTAAGATGGAGAGAGAAACTAGGTAAAGGAAATGTAAGAACAGCAGAGATAGCA	540
Db	1166	GGAGGTTAAGATGGAGAGAGAAACTAGGTAAAGGAAATGTAAGAACAGCAGAGATAGCA	1225
OY	541	GCCTGAAGAAAGTTATGGATGTGCTAGAGATGTTGAAATGTCAAACATGTATTTGCTGA	600
Db	1226	GCCTGAAGAAAGTTATGGATGTGCTAGAGATGTTGAAATGTCAAACATGTATTTGCTGA	1285
OY	601	CCAGAGGTTAATGGAAACCTAATGAGTTGAAATGTAGAAACCTCAGAAATGAAAGCTAG	660
Db	1286	CCAGAGGTTAATGGAAACCTAATGAGTTGAAATGTAGAAACCTCAGAAATGAAAGCTAG	1345
OY	661	CAAGAAATTGGAACACAGAAATGGAAATTTGAAATTGAGCCAGATTAAGAAATGTAAATCCCT	720
Db	1346	CAAGAAATTGGAACACAGAAATGGAAATTTGAAATTGAGCCAGATTAAGAAATGTAAATCCCT	1405
OY	721	TTCTCCTGGGAAAGAGAAATTCAGTCTTTAGACATGAAAAAGAGTCTGAGAAAAAGA	780
Db	1406	TTCTCCTGGGAAAGAGAAATTCAGTCTTTAGACATGAAAAAGAGTCTGAGAAAAAGA	1465
OY	781	AGAAAAAGAAATCTGAGCCCCCAACCTGAGCCTGGGCTCAACCTCAGCCTCAGTCCACCC	840
Db	1466	AGAAAAAGAAATCTGAGCCCCCAACCTGAGCCTGGGCTCAACCTCAGCCTCAGTCCACCC	1525
OY	841	CCAGCTTCAGCTTCATCCAGTCCCAACCCAGTACTCCAGTCCCAAGCCTCCCTCTCAGCC	900
Db	1526	CCAGCTTCAGCTTCATCCAGTCCCAACCCAGTACTCCAGTCCCAAGCCTCCCTCTCAGCC	1585
OY	901	TGAGGATTTGTCAATTAAGCTTTTAAAGCCAAACCCCAAGTTACTCAGAGGCAAGGGCA	960
Db	1586	TGAGGATTTGTCAATTAAGCTTTTAAAGCCAAACCCCAAGTTACTCAGAGGCAAGGGCA	1645
OY	961	TTTACTACTAGAGGAAGATTTTCTGTAGAGTCTGTAACAACTCAGTGAAGTACCACT	1020
Db	1646	TTTACTACTAGAGGAAGATTTTCTGTAGAGTCTGTAACAACTCAGTGAAGTACCACT	1705
OY	1021	AGAGCCAGTCTTGACAGTACATCCAGAGACCAAGAGCAAAACCAAACTAGAGACAGAG	1080
Db	1706	AGAGCCAGTCTTGACAGTACATCCAGAGACCAAGAGCAAAACCAAACTAGAGACAGAG	1765
OY	1081	TAGAGGTGACACTGGAATTAACCAAGCAAGACTAGAACTGCAAGCAGTAGCAGTAGAGAG	1140
Db	1766	TAGAGGTGACACTGGAATTAACCAAGCAAGACTAGAACTGCAAGCAGTAGCAGTAGAGAG	1825
OY	1141	TTCTAGTAGCAGTTCAACAGTAGACAGCACTGGAAGTAGTTCCAGACATGGAAAGTAGAG	1200
Db	1826	TTCTAGTAGCAGTTCAACAGTAGACAGCACTGGAAGTAGTTCCAGACATGGAAAGTAGAG	1885
OY	1201	CAGTGCAGTAGTTCAGATAGCAGCTTCAATACAATGTGCAGCAGACAGCAAGATATAGAG	1266
Db	1886	CAGTGCAGTAGTTCAGATAGCAGCTTCAATACAATGTGCAGCAGACAGCAAGATATAGAG	1945

Db 1290 CAAGAATTGGAACAGAAATGGAATTTGAAATGAGCCAGATGAAGATGTAATCCCT 1349
 Oy 721 TTCCCTGGGAAAGAAATGCAATGCTTTAGACATGAAAGAGAGCTGAGGAAAAAGA 780
 Db 1350 TTCTCTGGGAAAGAAATGCAATGCTTTAGACATGAAAGAGAGCTGAGGAAAAAGA 1409
 Oy 781 AGAAAAAGAAATGAGAGCCCAACCTGAGCTGTGGCTCAACCTGAGCTGAGCTGAGCC 840
 Db 1410 AGAAAAAGAAATGAGAGCCCAACCTGAGCTGTGGCTCAACCTGAGCTGAGCTGAGCC 1469
 Oy 841 CCAGCTTACCTTCAATCCAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCC 900
 Db 1470 CCAGCTTACCTTCAATCCAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCC 1529
 Oy 901 TGAGAGTTTGTCAATTTAGCTTTTACAGCCCAACCCCAAGTTACTGAGAGCAAGGCA 960
 Db 1530 TGAGAGTTTGTCAATTTAGCTTTTACAGCCCAACCCCAAGTTACTGAGAGCAAGGCA 1589
 Oy 961 TTTTCTACCTGAGAGGAAATTTCTCTGATAGCTGTAATCTACTAGTACCTAGT 1020
 Db 1590 TTTTCTACCTGAGAGGAAATTTCTCTGATAGCTGTAATCTACTAGTACCTAGT 1649
 Oy 1021 AGAGCCAGCTTTCAGAGTACATCCAGAGAGCAAGAGCAAAACCAAACTAGAGCAGAG 1080
 Db 1650 AGAGCCAGCTTTCAGAGTACATCCAGAGAGCAAGAGCAAAACCAAACTAGAGCAGAG 1709
 Oy 1081 TAGAGCTGAGCTGAGAAATTAACCAAGAGAGTGAAGTGAAGCAGTACAGTACAG 1140
 Db 1710 TAGAGCTGAGCTGAGAAATTAACCAAGAGAGTGAAGTGAAGCAGTACAGTACAG 1769
 Oy 1141 TTTAGTAGAGCTTTCAGAGTACATCCAGAGAGTGAAGTGAAGTGAAGTGAAGT 1200
 Db 1770 TTTAGTAGAGCTTTCAGAGTACATCCAGAGAGTGAAGTGAAGTGAAGTGAAGT 1829
 Oy 1201 CAGTCCAGAGTACCTGAGAGTACAGTACAGTACAGTACAGTACAGTACAGTACAG 1260
 Db 1830 CAGTCCAGAGTACCTGAGAGTACAGTACAGTACAGTACAGTACAGTACAGTACAG 1889
 Oy 1261 CAGTCCAGAGTACCTGAGAGTACAGTACAGTACAGTACAGTACAGTACAGTACAG 1320
 Db 1890 CAGTCCAGAGTACCTGAGAGTACAGTACAGTACAGTACAGTACAGTACAGTACAG 1949
 Oy 1321 AGATAGAAAGCAGAGAGAGGCTGATCGAGAGAGAGAGGATACCTGAGAGTACAG 1380
 Db 1950 AGATAGAAAGCAGAGAGAGGCTGATCGAGAGAGAGAGGATACCTGAGAGTACAG 2009
 Oy 1381 AGTCCAGAAATCTTCAAAAGAGTGTGATAGATACAAAGAGATCAAAAGATTAAGAA 1440
 Db 2010 AGTCCAGAAATCTTCAAAAGAGTGTGATAGATACAAAGAGATCAAAAGATTAAGAA 2069
 Oy 1441 TTCCCGGTCGAGAGAAAGAGTGTATATCAGAGAGTACGATCAGGCAAAAGATCTTC 1500
 Db 2070 TTCCCGGTCGAGAGAAAGAGTGTATATCAGAGAGTACGATCAGGCAAAAGATCTTC 2129
 Oy 1501 AAGAAAGTGAAGAGAGAGGCTGATAGAGAGAGAGAGAGGCTGATAGAGAGAGAGG 1560
 Db 2130 AAGAAAGTGAAGAGAGAGGCTGATAGAGAGAGAGAGAGGCTGATAGAGAGAGAGG 2189
 Oy 1561 AGGCTTCTTACGAGATCTTTCAGAGAGAGATTTCTGATAGAGAGAGATTTACCTTCC 1620
 Db 2190 AGGCTTCTTACGAGATCTTTCAGAGAGAGATTTCTGATAGAGAGAGATTTACCTTCC 2249
 Oy 1621 TTGTAAGAGAGAGAGGCTGATAGAGAGAGAGAGAGGCTGATAGAGAGAGAGG 1680
 Db 2250 TTGTAAGAGAGAGAGGCTGATAGAGAGAGAGAGAGGCTGATAGAGAGAGAGG 2309
 Oy 1681 GACTGTGTTTACGAGATCTTTCAGAGAGAGATTTCTGATAGAGAGAGATTTATATC 1740
 Db 2310 GACTGTGTTTACGAGATCTTTCAGAGAGAGATTTCTGATAGAGAGATTTATATC 2369

RESULT 5
 HS077718

LOCUS HS077718 2617 bp mRNA Linear PRI 01-MAR-1997
 DEFINITION Human desmosome associated protein p19in mRNA, complete cds.
 ACCESSION U77718
 VERSION 077718.1 GI:1684846
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homini; Homo.
 REFERENCE
 Ouyang, P. and Sugrue, S.P.
 Characterization of p19in, a novel protein associated with the
 desmosome-intermediate filament complex
 J. Cell Biol. 135 (4), 1027-1042 (1996)
 JOURNAL MEDLINE 97081102
 REFERENCE 2 (bases 1 to 2617)
 Sugrue, S.P. and Ouyang, P.
 Direct Submission
 Submitted (08-NOV-1996) Anatomy and Cell Biology, University of
 Florida, 1600 SW Archer Road, Gainesville, FL 32610, USA
 FEATURES
 source
 CDS
 31..2262
 /note="desmosome associated protein; phosphoprotein"
 /protein_start=1
 /product="p19in"
 /protein_id="AAB48304.1"
 /db_xref="GI:1684847"
 /translation="MAVAVRTLODELAKESLNVNDENIRKLGRPNDRPIQARL
 IALSGRGGRGSLILRRGFSDSGSPAKQNDLEGAVSRIGERRTRREROESDPE
 DDVVKFPAQSSVAVTSKRTNRDLQDNMDKGRKNRRIFGMLGTLQFQEST
 VATERONRRQLEQKLEVOAEERQVENERELEFEERRARQTELRILEQVLEALQ
 EEMNEHNAIKIRYIRKTKPLFLYIPGRCPATQKLIIESQRKNALEDRVIEAEG
 IKMEAPRPRQMKERHQVYVNEHSEKQAEQKQAEELVETGNOHNDVIEAEG
 EEEKEIGIVHSDAEKEDEEKOQEMAEKLEETEVESEKQDSEPEVDVLEAY
 ENVKHYIAQVEHETNRYVESVPSNEAEKLEPEMERIEPDKECKSLGKENVSA
 LMEKESDEKESEPEPEVAPQAOQLOPOLQOSPEPOLOPPAPLOPOLQOS
 QVLOQSOCHAVIVQSHRPSQPELSTLAIVLOPTVOYQENGHLPKRPDPVSYKITEY
 PDVPLVTHPESEETNRKSRGRNRKNTTKSRSSSSSSSSSSSSSSSSSSSSSSSSSSSS
 GSSSSSSSSSSSSSTSGSS
 TSGLEKSHRSKSGSSSDTKSGSKDNRSRDRKRSISESSRSGKRSRSEDRKSDRKD
 KRR"
 BASE COUNT 911 a 471 c 686 g 549 t
 ORIGIN
 Query Match 91.5%; Score 1646.6; DB 9; Length 2617;
 Best Local Similarity 93.6%; Pred. No. 3; 1e-309;
 Matches 1761; Conservative 0; Mismatches 39; Indels 81; Gaps 1;
 Oy 1 ACAGACAGACTGGGCTTTTGGAGAGAGTTGAGCTTGGCCAGCTGCAAGAGAAATG 60
 Db 633 ACAGACAGAACTGGGCTTTTGGAGAGAGTTGAGCTTGGCCAGCTGCAAGAGAAATG 692
 Oy 61 GAATGAACATATATCCCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 120
 Db 693 GAATGAACATATATCCCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 752
 Oy 121 TTATATTCCTGGAAGAGTGTCCAGTACCAAAATTAATTAATTAATTAATTAATTAATTA 180
 Db 753 TTATATTCCTGGAAGAGTGTCCAGTACCAAAATTAATTAATTAATTAATTAATTAATTA 812
 Oy 181 AATGAACGCTTTATTTGAAGTAGAGCCATGCAATTTGAGAGCAAAATTAATTAATTAATGA 240
 Db 813 AATGAACGCTTTATTTGAAGTAGAGCCATGCAATTTGAGAGCAAAATTAATTAATTAATGA 872
 Oy 241 GCGTAGGCTTAGAAGACAAATCAATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
 Db 873 GCGTAGGCTTAGAAGACAAATCAATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 932
 Oy 301 ACAGAAAGGCGAACAAGAGAGGTAAGTGGCTCAGCGAGAGAGAGAGAGAGAGAGAGAG 360

```

Db 933 ACACAAAGCGGAAACAGAGAGGTAAGTGCTCAGCCAGAGGAGAGTGGTGAGAC 992
Oy 361 AGGTAATACGACCAATGATGTAAGAAATAGAGNAGCAGAGAGAGAGAGAGAAAT 420
Db 993 AGGTAAACGACCAATGATGTTGAAATAGAGAGCAGAGAGAGAGAGAGAGAAAT 1052
Oy 421 AGGATTTGTCATGATGATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
Db 1053 AGGATTTGTCATGATGATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1112
Oy 481 GAGGTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 540
Db 1113 GAGGTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1172
Oy 541 GCGTGAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600
Db 1173 GCGTGAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1232
Oy 601 CCGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 660
Db 1233 CCGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1292
Oy 661 CAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 720
Db 1293 CAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1352
Oy 721 TTTCTCGGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATG 780
Db 1353 TTTCTCGGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATG 1412
Oy 781 AGAAAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 840
Db 1413 AGAAAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1472
Oy 841 CCAAGTTCAGCTTCAATCCAGTC----- 864
Db 1473 CCAAGTTCAGCTTCAATCCAGTC----- 1532
Oy 865 -----CCAACCTACTCCCA 879
Db 1533 ACCTCAGCTTCAATCCAGTC----- 1592
Oy 880 GTCCCAAGCTTCAATCCAGTC----- 939
Db 1593 GTCCCAAGCTTCAATCCAGTC----- 1652
Oy 940 AGTTACTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 999
Db 1653 AGTTACTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1712
Oy 1000 AAAAATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1059
Db 1713 AAAAATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1772
Oy 1060 AACCAAAATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1119
Db 1773 AACCAAAATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1832
Oy 1120 TCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1179
Db 1833 TCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1892
Oy 1180 TCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1239
Db 1893 TCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1952
Oy 1240 CAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1299
Db 1953 CAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2012
Oy 1300 TAGGGGTGGGAGATATATAGATATAGAAAGCAGAAAGAGAGAGAGAGAG 1359

```

```

Db 2013 TAGGGGTGGGAGATATATAGATATAGAAAGCAGAAAGAGAGAGAGAGAG 2072
Oy 1360 GGATACCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1419
Db 2073 GGATACCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2132
Oy 1420 AAAAGATCAAAAGATCAAAAGATCAAAAGATCAAAAGATCAAAAGATCAAA 1479
Db 2133 AAAAGATCAAAAGATCAAAAGATCAAAAGATCAAAAGATCAAAAGATCAAA 2192
Oy 1480 TCGATGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1539
Db 2193 TCGATGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2252
Oy 1540 AAGGCTTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1599
Db 2253 AAGGCTTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2312
Oy 1600 ATAAAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1659
Db 2313 ATAAAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2372
Oy 1660 AAAATCTTTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1719
Db 2373 AAAATCTTTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2432
Oy 1720 ATTTTGAAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1779
Db 2433 ATTTTGAAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2492
Oy 1780 AATAAAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1800
Db 2493 AATAAAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2513

RESULT 6
CNS01DT4 196319 bp DNA linear PRI 29-APR-2001
LOCUS Human chromosome 14 DNA sequence BAC R-407N17 of library RPI-11
DEFINITION from chromosome 14 of Homo sapiens (human), complete sequence.
ACCESSION AL132639 GI:13897473
VERSION AL132639.4
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 196319)
AUTHORS Heilig,R., Petit,J.L., Vico,V., Dasilva,C., Robert,C., Wincker,P.,
Brottier,P., Catolico,L., Barbe,Y., Pellelet,E., Artiguenave,F.,
Levy,M., Eckenberg,R., Bruls,T., Deberardinis,V., Cruaud,C.,
Gyapay,G., Saurin,W. and Weissensbach,J.
Sequencing of the human chromosome 14
unpublished
2 (bases 1 to 196319)
Genoscope.
REFERENCE Direct Submission
AUTHORS Genoscope - Centre National de Sequencage :
JOURNAL Submitted (28-APR-2001) Genoscope - Centre National de Sequencage :
BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
- Apr 30, 2001 this sequence version replaced gi:12001725.

COMMENT --Genome Center
Center: Genoscope / Centre National de Sequencage
Center code: GS
Web site: http://www.genoscope.cns.fr/
Contact: seqref@genoscope.cns.fr

The following BAC sequence is oriented from the T7 to the SP6 end.
Upstream BAC (overlapping the T7 end) : R-545M17
Downstream BAC (overlapping the SP6 end) : R-647020 (AC-AL157791)
----- Summary Statistics -----
Assembly program: Phrap: version 2.0
Quality coverage: 6.71x in Q20 bases; sum-of-contigs

```

Overall quality chart :
Range : bases
0 :
1 - 9 : 5
10 - 19 : 19
20 - 29 : 19
30 - 39 : 254
40 - 49 : 4102
50 - 59 : 9179
60 - 69 : 14852
70 - 79 : 33313
80 - 89 : 69330
90 - 99 : 65265

Percentage of bases with a quality value >= 40 : 99 %.

FEATURES
source

1. 196319
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/clone="R-407N17"
/clone_1lb="R-PC1-11"
24883..25024
/note="matching EMBL:R15752
Rdb:RH4450
dbSTS:STS37533
Identified using the e-PCR software (G. Schuler)*
24943..25069
/note="matching EMBL:R15752
Rdb:RH53971
dbSTS:STS9647
Identified using the e-PCR software (G. Schuler)*
25391..25495
/note="matching EMBL:D20234
Rdb:RH44089
Rdb:RH160
dbSTS:STS18010
Identified using the e-PCR software (G. Schuler)*
59383..59517
/note="matching EMBL:G19682
Rdb:RH66850
dbSTS:STS46779
Identified using the e-PCR software (G. Schuler)*
141703..141878
/note="matching EMBL:T59353
Rdb:RH75313
dbSTS:STS52401
Identified using the e-PCR software (G. Schuler)*
BASE COUNT 55394 a 37759 c 40274 g 62892 t
ORIGIN

Query Match 89.1%; Score 1603.4; DB 9; Length 196319;
Best Local Similarity 99.6%; Pred. No. 8.5e-301;
Matches 1607; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

188 GCCTTATTAGAGTACCATCATGATTTGCAGAACAAATAATAATATGAGGCTAGG 247
|||||
Db 23332 GCCTTATTAGAGTACCATCATGATTTGCAGAACAAATAATAATATGAGGCTAGG 23391
248 CCTAGAACATCATGAG 307
|||||
Db 23392 CCTAGAACATCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 23451
308 GCGGAAACAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 367
|||||
Db 23452 GCGGAAACAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 23511
368 CAGCACAATGATGAGAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 427
|||||
Db 23512 CAGCACAATGATGAGAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 23571

428 GTTCATAGTATGCGAGAGAAAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 487
|||||
Db 23572 GTTCATAGTATGCGAGAGAAAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 23631
488 AAGATGAGAGAGAGAACTGAGTAAAGGAAAGTGAAGAGAGAGAGAGAGAGAGAGAGT 547
|||||
Db 23632 AAGATGAGAGAGAGAACTGAGTAAAGGAAAGTGAAGAGAGAGAGAGAGAGAGAGT 23691
548 GAAAGTTATGATGCTGTGAGATGAGTGAAGATGTCAAACATGTAATTTGCTGACACGAG 607
|||||
Db 23692 GAAAGTTATGATGCTGTGAGATGAGTGAAGATGTCAAACATGTAATTTGCTGACACGAG 23751
608 GTAAATGAAACTAATCGAGATTGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 667
|||||
Db 23752 GTAAATGAAACTAATCGAGATTGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 23811
668 TTGGAACCGAAATGGAATTTGAAATGAGCCAGATTAAGAAATGTAATTCCTTCTCT 727
|||||
Db 23812 TTGGAACCGAAATGGAATTTGAAATGAGCCAGATTAAGAAATGTAATTCCTTCTCT 23871
728 GCGAAAGAGAAATGTCAGTCTTTAGACATGGAAGAGAGTGTGAGAGAGAGAGAGAGAG 787
|||||
Db 23872 GCGAAAGAGAAATGTCAGTCTTTAGACATGGAAGAGAGTGTGAGAGAGAGAGAGAGAG 23931
788 GAATCTGAGCCCCAACCTGAGCCTGTGCTCAACCTCAACCTCAACCTCAACCTCAACCT 847
|||||
Db 23932 GAATCTGAGCCCCAACCTGAGCCTGTGCTCAACCTCAACCTCAACCTCAACCTCAACCT 23991
848 CAGCTTCATCCAGTCCCAACAGTACTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGT 907
|||||
Db 23992 CAGCTTCATCCAGTCCCAACAGTACTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGT 24051
908 TTGTCTATGCTGTTTACAGCCAAACCCCAAGTTACTCAGAGAGAGAGAGAGAGAGT 967
|||||
Db 24052 TTGTCTATGCTGTTTACAGCCAAACCCCAAGTTACTCAGAGAGAGAGAGAGAGAGT 24111
968 CCGAGAGAGAGAGATTTTCTGTAGAGTCTGTAAGAACTCAGTGAAGTCCAGTGAAGAGCA 1027
|||||
Db 24112 CCGAGAGAGAGAGATTTTCTGTAGAGTCTGTAAGAACTCAGTGAAGTCCAGTGAAGAGCA 24171
1028 GTCTTGACAGTACATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 1087
|||||
Db 24172 GTCTTGACAGTACATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 24231
1088 CGAGCTAGAAATTAACAAAGCAAGAGTGAAGTGCAGAGAGAGAGAGAGAGAGTCTAGT 1147
|||||
Db 24232 CGAGCTAGAAATTAACAAAGCAAGAGTGAAGTGCAGAGAGAGAGAGAGAGAGTCTAGT 24291
1148 AGCAGTTCACAGAGTACAGCAGAGTGAAGTTCAGCAGTGAAGTGAAGTGAAGTGAAGT 1207
|||||
Db 24292 AGCAGTTCACAGAGTACAGCAGAGTGAAGTTCAGCAGTGAAGTGAAGTGAAGTGAAGT 24351
1208 AGTAGTTCAGTAGAGCTCCAGTACAGTGAAGTGGCAGAGAGAGAGAGAGAGTGAAGTGAAG 1267
|||||
Db 24352 AGTAGTTCAGTAGAGCTCCAGTACAGTGAAGTGGCAGAGAGAGAGAGAGAGTGAAGTGAAG 24411
24412 ACTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAG 24471
1328 AAGCACAAGAGAGCTGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGTAC 1387
|||||
Db 24472 AAGCACAAGAGAGCTGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGTAC 24531
1388 AATCTTCAAAAGAGTGTGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAG 1447
|||||
Db 24532 AATCTTCAAAAGAGTGTGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAG 24591
1448 TCCGACAGAAAGAGTCTATATCAGAGAGTATGATTCAGGCAAAAGATCTTCAAGAAAGT 1507
|||||
Db 24592 TCCGACAGAAAGAGTCTATATCAGAGAGTATGATTCAGGCAAAAGATCTTCAAGAAAGT 24651
1508 GAAAGAGCCCGAAATCAGACAG 1567

DB 3365 TTCTGTAGAAATGTGTAATAAATCACTAGCTAGGTTACCGACAGAGCCAGTCTTGACGGTACATTT 3424
OY 1044 CAGAGAGCAAGAGCAAAACCCAAAACCTAGAGCAGCAAGTACAGCTCGACGTAGAAATATAA 1103
 |||||
DB 3425 CAGAGAGCAAGTACGAAACCCAAAACCTAGAGCAGCAAGTACAGCTCGACGTAGAAATATAA 3484
 |||||
OY 1104 CAGCAAGAGTAGAAGTCGAAAGCAGTAGAGCAGTACAGTCTTCTAGTAGCAGTTCACACAGTA 1163
 |||||
DB 3485 CAGCAAGAGTAGAAGTCGAAAGCAGTAGAGCAGTACAGTCTTCTAGTAGCAGTTCACACAGTA 3544
 |||||
OY 1164 GCAGCAAGTAGAAGTACGTCACAGCAGTGGAAAGTACAGCTCGCAGTAGTCTCCAGTAGCA 1223
 |||||
DB 3545 GCAGCAAGTAGAAGTACGTCACAGCAGTGGCAAGTACAGTACAGTACGTCGACCAAGT---AGCAGCT 3601
 |||||
OY 1224 GCTCCAGTCAAGTAGCGCAGCAGCAGCAGCAAGTATAGTACAGTACAGTACAGTACAGTACAGT 1283
 |||||
DB 3602 CAGGTAGTCAAGTACGTCACAGCAGCAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGT 3661
 |||||
OY 1284 AGAGTAGAAGTCCGAGTAGTGGGGGCTCGGGGACATTAATAGAGATAGAAAGCAGCAAGAGAGCG 1343
 |||||
DB 3662 AGAGCAAGTCCGAGTAGTGGGGGCTCGGGGACATTAACAGAGATAGAAAGCAGCAAGAGAGAGCG 3721
 |||||
OY 1344 TGCATCGAAGAGAGAGGATACCTCAGAGCTAGAGAAAGCAGCAAGTCTTCCAAAGAGCTG 1403
 |||||
DB 3722 TGCATCGAAGAGAGGATACCTCAGAGCTAGAGAAAGCAGCAAGTCTTCCAAAGAGCTG 3781
 |||||
OY 1404 GTGGTAGTAGAGATACAAAGGA 1426
 |||||
DB 3782 GTAGTAGTAGAGATACAAAGGA 3804
 |||||

RESULT 12
PIGSEBNU 1109 bp mRNA linear MAM 07-MAR-1995
LOCUS PIGSEBNU
DEFINITION Pig neutrophil protein mRNA, 3' end.
ACCESSION M55701
VERSION M55701.1 GI:164672
KEYWORDS neutrophil protein.
SOURCE Pig neutrophil, cDNA to mRNA.
ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
 Bellavite, P., Bazzone, F., Cassatella, M.A., Hunter, K.J. and
 Bannister, J.V.
 Isolation and characterization of a cDNA clone for a novel
 serine-rich neutrophil protein.
 Biochem. Biophys. Res. Commun. 170 (2), 915-922 (1990)
 90343818
 Location/Qualifiers
 1..1109

BASE COUNT	360 a	231 c	269 g	229 t	
ORIGIN					
Query Match	44.8%	Score 806.8:	DB 4:	Length 1109:	
Best Local Similarity	88.2%	Fred. No. 1.7e-146:			
Matches 975:	Conservative	0:	Mismatches 82:	Indels 49:	Gaps 7

Oy	736	GAGTGTGAGGCTTTTGACATGGAAAAAGAGCTGAGAAAAACAAACAAAAAGATCTGA	795
Db	1	GAATTCCTAGTGCCTCTGAAATGTGAAAAATGAGCTTAGGAAAAAAGAAAAAGATCCGA	60
Oy	796	GCCCCAACCTGAGCCCTGTGGCTCAACCTCAGCCCTCATGTCAGGCCAGC-----	845
Oy	846	-----TTAGGCTTAATCCAGTCCCAACCAAG	873
Db	121	GCCACAGCCACAGCACAGGCCCCAGACCCAGGCCCAAGTGCACAGTCCCAACCAAGC	180
Oy	874	ACTCCAGTCCAGGCTCCCTCTCTCAGCCCTGAGGATTTGTCATTAGCTGTTTACAGCCAAC	933
Db	181	ACTCCAGCCCAAGCTCTCTCTCAGCCCTGAGACTATACCCCTTAGCTGTTTACAGCCACC	240
Oy	934	ACCCCAAGTTACTCAGAGCAAGGCAATTACTACCTGAGAGAAAGATTTTCTGTAGA	993
Db	241	ACCTCAGGCTATTCCAGAGCAAGGGATTTACCCTCTGAGAAAGGTTTCTGTAGA	300
Oy	994	GTCGTAAAACTCCTAGGTACCACTAGAGCAAGCTCTGACAGTACATCCAGAGACAA	1053
Db	301	ATCTCCAAAACTGACTAGGTGACATGAGGCCAGTCTTGACAGTACATCCAGAGACAA	360
Oy	1054	GAGCAAAACCAAACTAGAGACAGAAAGTAGAGGTGAGCTAGAAATTAACACCAAGAG	1113
Db	361	GAGCAAAACCAAACTAGAGACAGAAAGTAGAGGTGAGCTAGAAATTAACACCAAGAG	420
Oy	1114	TAGAAGTGAAGCAGTAGCAGTAGCAGTTCTAGTACCACTTAACACAGTACAGCAGTGG	1173
Db	421	CAGAAGTGAAGCAGTAGCAGTAGCAGTTCTAGTACCACTTAACACAGTACAGCAGTGG	480
Oy	1174	AAGTAGTCCAGCAGTGGAAAGTAGTAGCAGTCGACAGTACTTCCAGTACAGCTCCAGTAC	1233
Db	481	AAGCAGTCTCAGAGTGGCAGTAGTAGTAGTCCAGTAGTACTTCCAGTACAGCTCC	539
Oy	1234	AAGTGCAGCAGCAGCAGCAGCAGTGAAGCAGTGAAGCAGTGAAGCAGTGAAGCAGTGAAG	1292
Db	540	AAGTGCAGCAGCAGCAGCAGTGAAGCAGTGAAGCAGTGAAGCAGTGAAGCAGTGAAGCAGTGAAG	599
Oy	1293	GTCGAGTAGAGGCGTTCGGGGACATTAATAGAGATGAAGAACACAGAAAGAGGCGTGGATCGA	1352
Db	600	GTCGAGTAGAGGCGCGGGACATTAACAGAGATGAAGAACACAGAGGAGGAGTGGATCGGA	659
Oy	1353	AGAGAAAGGAGTACTTTCAGAGCAGTGAAGAAAGTCCAAATCTTCAAAAGGTGGTGTAGTA	1412
Db	660	ACCGGAGGAGTCTCTGAGACTAGAAAGAAAGTCCAAATCTTCAAAAGGTGGTGTAGTA	719
Oy	1413	GAGATACAAAAGGATCAAAAGGATTAAGAAATCCCGGTCCGACAGAAAGAGGTCTATACAG	1472
Db	720	GAGATACAAAAGGATCAAAAGGATTAAGAAATCCCGGTCTGACAGAAAGGTCTATATACAG	779
Oy	1473	AGAGTAGTCGATCAGGCAAAAGATCTTCAAGAAAGTGAAGAGCCCGGAAATTCAGACAGGA	1532
Db	780	AGAGTAGTCGATCAGGCAAAAGATCTTCAAGAAAGTGAAGAGCCCGGAAATTCAGACAGGA	839
Oy	1533	AAGACAAAAGGGTAAATGGAAGAGCAGGCTTTCTTAGACCATTCTTGCAGCAGAGA	1592
Db	840	AAGACAAAAGGGCTTAATGGAAGAGCAGGCTTTCTTAGACCATTCTTGCAGCAGAGA	899
Oy	1593	TTTCTTGATTA--AAAGAGTATTCCTTCTGTGAAGAGAGTCTGCTTAAGAATTGC	1650
Db	900	TTTCTTGATGAGTAATAAGAGATTACCTTCTGTTGAAGAGAAATGCTGCTTAAGAATTGC	959
Oy	1651	ATGCTGTAAAAATCTTTTGAAAAATACAGAGCTTTGTTTACCGAGACATCTTGTACT	1710
Db	960	ATGCTGT--AAAAATTTTGTGAAAAATACAGACTGTTTGTTCACAGAC--TTCTTGAC-	1016
Oy	1711	TTTTGCATAAATTTGTAAGAGTTATTATCAAAATTAATGAGTTCCAAATATGTAAA	1770
Db	1017	TTTTGCATAAATTTGTAAGAGTTATTATCAAAATTAATGAGTTCCAAATATGTGTAAA	1076
Oy	1771	AATGATAAATATAAAAAAGATTAC	1796

Db 1077 AATATATATATATATATATATAC 1102

RESULT 13

AC100145

Mus musculus clone RP23-43A22, LOW-PASS SEQUENCE SAMPLING.

AC100145.1 GI:17047511

HTG: HTGS_PHASE0.

house mouse.

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AC100145 62685 bp DNA 11near HTG 22-NOV-2001
Mus musculus clone RP23-43A22, LOW-PASS SEQUENCE SAMPLING.
AC100145.1 GI:17047511
HTG: HTGS_PHASE0.
house mouse.
Mus musculus
Eukaryota: Chordata: Craniata: Vertebrata: Euteleostomi:
Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
1 (bases 1 to 62685)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Unpublished
2 (bases 1 to 62685)
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Bouhgalter, B.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Choe, P., Colangelo, M., Collins, S., Collamore, A., Cook, A.,
Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S.,
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kells, C., Labrecque, K.,
Lamarez, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
MacLean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meldrum, J.,
Meneus, L., Mihova, T., Mlenka, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, D., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retter, R., Rieback, M., Riley, R., Risse, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R.,
Seaman, S., Severy, P., Spencer, B., Stange, Thoman, N., Stojanovic, N.,
Strauss, N., Sudamanian, A., Talamas, J., Testaye, S., Theodores, J.,
Topham, R., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zaloun, J., Zembek, L., Zimmer, A. and Zody, M.
Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
Center: Whitehead Institute/MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: L14033
Center clone name: 43_A_22

* NOTE: This record contains 77 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

2276 2475: gap of 100 bp
2476 3179: contig of 704 bp in length
3180 3279: gap of 100 bp
3280 3979: contig of 700 bp in length
3980 4079: gap of 100 bp
4080 4795: contig of 716 bp in length
4796 4895: gap of 100 bp
4896 5610: contig of 715 bp in length
5611 5710: gap of 100 bp
5711 6444: contig of 734 bp in length
6445 6544: gap of 100 bp
6545 7266: contig of 722 bp in length
7267 7366: gap of 100 bp
7367 8073: contig of 707 bp in length
8074 8173: gap of 100 bp
8174 8903: contig of 730 bp in length
8904 9003: gap of 100 bp
9004 9713: contig of 710 bp in length
9714 9813: gap of 100 bp
9814 10518: contig of 705 bp in length
10519 10618: gap of 100 bp
10619 11347: contig of 729 bp in length
11348 11447: gap of 100 bp
11448 12155: contig of 708 bp in length
12156 12255: gap of 100 bp
12256 12961: contig of 706 bp in length
12962 13061: gap of 100 bp
13062 13783: contig of 722 bp in length
13784 13883: gap of 100 bp
13884 14619: contig of 736 bp in length
14620 14719: gap of 100 bp
14720 15439: contig of 720 bp in length
15440 15539: gap of 100 bp
15540 16275: contig of 736 bp in length
16276 16375: gap of 100 bp
16376 17102: contig of 727 bp in length
17103 17202: gap of 100 bp
17203 17914: contig of 712 bp in length
17915 18014: gap of 100 bp
18015 18714: contig of 700 bp in length
18715 18814: gap of 100 bp
18815 19530: contig of 716 bp in length
19531 19630: gap of 100 bp
19631 20353: contig of 723 bp in length
20354 20453: gap of 100 bp
20454 21154: contig of 701 bp in length
21155 21254: gap of 100 bp
21255 21967: contig of 713 bp in length
21968 22067: gap of 100 bp
22068 22818: contig of 751 bp in length
22819 22918: gap of 100 bp
22919 23649: contig of 731 bp in length
23650 23749: gap of 100 bp
23750 24449: contig of 700 bp in length
24450 24549: gap of 100 bp
24550 25278: contig of 729 bp in length
25279 25378: gap of 100 bp
25379 26078: contig of 700 bp in length
26079 26178: gap of 100 bp
26179 26908: contig of 730 bp in length
26909 27008: gap of 100 bp
27009 27720: contig of 712 bp in length
27721 27820: gap of 100 bp
27821 28521: contig of 701 bp in length
28522 28621: gap of 100 bp
28622 29339: contig of 718 bp in length
29340 29439: gap of 100 bp
29440 30137: contig of 698 bp in length
30138 30237: gap of 100 bp
30238 30933: contig of 696 bp in length
30934 31033: gap of 100 bp
31034 31743: contig of 710 bp in length
31744 31843: gap of 100 bp

TITLE	JOURNAL	REFERENCE	AUTHORS	TITLE	JOURNAL
Direct Submission	Unpublished	2 (bases 1 to 49360)	Morley R.C.	Direct Submission	Submitted (07-FEB-2002)
					Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
					----- Genome Center -----


```

OY 1012 GGTACACAGTAGCCAGTCTTGACAGTACATCCAGAGACAGCAAAACCAAACTAG 1071
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1583 GGCACAGATATATATATGCCCCAAAACCAACCTCCCTCCATTAATAA----- 1631
OY 1072 GAGCAGAGTAGAGCTGAGCTAGAAATATAAACAAGCAGAGTAGAGTAGAGTAG 1131
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1632 ----AGGAGTAGGAGCAGAGGAGAAATATAAACAAGATGGGTAAAGCAGAGTAGTAG 1687
OY 1132 CAGTAGCAGTCTAGTACAGTTCACACAGCAGAGTAGGAGTAGTCCACAGTAGG 1191
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1688 TACGACAGTAGTTCGAGTAGTACTTCTAGCTTAGTACTAGTGGAGCAGTCTAGT----- 1742
OY 1192 AAGTAGTAGCAGTGCAGTAGTCTCCAGTAGCAGTCCAGTAGGAGCAGCAGCAG 1251
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1743 -AGTAGCAGTAGCCAAAGCAGAGTTCACAGCAGCAGTAGCAGCAGTGGAAAGCAGCAGCCG 1801
OY 1252 AGATAGTAGCAGTACAGTCTAGTACTAGTACAGAGTAGAGTAGAGTAGAGGAGTCCGGG 1311
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1802 AGACAGCAGCAGTAGTAGCAGCAGCAGTAGTAGGAAAGTGGAAAGCCGACAGCAGAGTCTGTGG 1861
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1312 ACATATATAGATAGAAAGCAGCAGAGAGGCGGTGATCGAAGAGAGGATCTTCAGG 1371
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1862 GCATGCCAGAGATGAAAAAGCAGAGAGATGGCGAGCAGCAGAGAGAGATGGAGCAGG 1921
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1372 ACTGAAAAGAGTACAAATCTTCAAAAGTGGTGTAGTAGATACAAAAGATCAAA 1431
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1922 TGTAGAGAGAGTCAAGAGTCTTCAAAAGGCAAGT---GCAGAGACTCCAAAGGATCCAA 1978
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1432 GGATAGAGATTCGCGGTCGACAGAAAGAGTCTATATCAGAGAGTAGTGCATCAGGCCAA 1491
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1979 GGACAAAAGTTCAGAGTCTGACAGAGAGAGACTACTTTCAGAGAGTAGTGGTCAAGGCCAA 2038
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1492 AAGATCTTCAAGAGAGAGAGCCGCAAAATCAGACAGCAAAAGCAGCGTTAA 1549
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2039 GCGACCTCTCAAGAAATGAAAGAGACCGTAAATCAGACAGCAAAAGACGTTAA 2096
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: October 16, 2002, 14:08:41
 Job time : 3784 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 16, 2002, 12:50:13 ; Search time 65 Seconds
(without alignments)
6802.159 Million cell updates/sec

Title: US-09-822-849A-1
Perfect score: 1800
Sequence: 1 acagacagactgcgcgtt.....ataaaaaagattacatcc 1800

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 38353 seqs, 122816752 residues
Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /cgn2_6/pdata/2/lna/5A.COMB.seq:*
2: /cgn2_6/pdata/2/lna/5B.COMB.seq:*
3: /cgn2_6/pdata/2/lna/6A.COMB.seq:*
4: /cgn2_6/pdata/2/lna/6B.COMB.seq:*
5: /cgn2_6/pdata/2/lna/PCTUS.COMB.seq:*
6: /cgn2_6/pdata/2/lna/Dackfilseq1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1730.4	96.1	2369	4	US-08-910-925-2
2	125.6	7.0	7218	1	US-08-232-463-14
3	121.2	6.7	3469	2	US-08-728-323A-1
4	121.2	6.7	32207	2	US-08-770-379-20
5	121.2	6.7	32207	4	US-08-757-669A-20
6	121.2	6.7	32207	4	US-09-230-371A-20
7	102.6	5.7	397	3	US-09-253-691-3
8	99.2	5.5	16442	3	US-08-781-891-208
9	97.2	5.4	3337	2	US-08-072-610-1
10	97.2	5.4	3337	2	US-08-719-822B-1
11	97.2	5.4	3337	4	US-09-032-458-1
12	89.8	5.0	234	1	US-08-469-802B-3
13	89.8	5.0	234	1	US-08-267-803B-3
14	89.8	5.0	543	6	5273901-6
15	86.6	4.8	203	4	US-09-043-303-7
16	81.8	4.5	533	6	5482709-5
17	81.2	4.5	5511	3	US-08-928-361B-2
18	81.2	4.5	7334	3	US-08-928-361B-1
19	81	4.5	195	1	US-08-469-802B-2
20	81	4.5	195	2	US-08-267-803B-2
21	80.6	4.5	51259	3	US-08-781-891-209
22	79.8	4.4	165	4	US-09-043-303-17
23	79.4	4.4	2580	3	US-09-050-863-2
24	79.4	4.4	2580	3	US-09-359-081-2
25	79.4	4.4	5452	2	US-09-130-114-1
26	79.4	4.4	9600	4	US-08-910-647-1
27	79.4	4.4	10596	1	US-07-884-811-15

28	79.4	4.4	10596	1	US-07-885-971-15	Sequence 15, Appl
29	79.4	4.4	10596	1	US-08-087-783A-15	Sequence 15, Appl
30	79.4	4.4	10596	1	US-08-194-088B-15	Sequence 15, Appl
31	79.4	4.4	10596	2	US-08-194-087-15	Sequence 15, Appl
32	79.4	4.4	10596	2	PCT-US93-04648-15	Sequence 15, Appl
33	79	4.4	477	4	US-09-135-994-1	Sequence 1, Appl
34	75.2	4.2	154	1	US-08-469-802B-6	Sequence 6, Appl
35	75.2	4.2	154	2	US-08-267-803B-6	Sequence 6, Appl
36	74.2	4.1	3376	3	US-08-320-559-29	Sequence 29, Appl
37	74.2	4.1	3376	3	US-08-545-860D-29	Sequence 29, Appl
38	74.2	4.1	3376	5	PCT-US94-04496-29	Sequence 5, Appl
39	74	4.1	171	1	US-08-469-802B-5	Sequence 5, Appl
40	74	4.1	171	2	US-08-267-803B-5	Sequence 5, Appl
41	72	4.0	168	1	US-08-469-802B-4	Sequence 4, Appl
42	72	4.0	168	2	US-08-267-803B-4	Sequence 4, Appl
43	71.8	4.0	1276	4	US-09-177-325-2	Sequence 2, Appl
44	71.8	4.0	1276	4	US-09-411-812A-2	Sequence 2, Appl
45	71.8	4.0	1276	4	US-09-590-113-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-910-925-2
Sequence 2, Application US/08910925
Patent No. 6162601
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Lal, Preeti
APPLICANT: Shah, Purvi
TITLE OF INVENTION: HUMAN PININ SPLICE VARIANT
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/910,925
FILING DATE: Herewith
CLASSIFICATION: 514
PRIOR APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0365 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2369 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: FIBROT01
CLONE: 53219
US-08-910-925-2
Query Match 96.1%; Score 1730.4; DB 4; Length 2369;
Best Local Similarity 99.7%; Pred. No. 0;

APPLICATION NUMBER: US/07/935, 113
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29, 768
REFERENCE/DOCKEY NUMBER: 30472/114 INDI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
DIENDATE SOURCE:
CLONE: PT3pt-F15
-08-232-463-14

[illegible]

RESULT 5

US-08-757-669A-20/c
; Sequence 20, Application US/08757669A
; Patent No. 6183751

GENERAL INFORMATION

APPLICANT: Chang, Yuan
 APPLICANT: Bohenzky, Roy A.
 APPLICANT: Russo, James J.
 APPLICANT: Edelman, Isidore S.
 TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S S
 TITLE OF INVENTION: SEQUENCES AND USES THEREOF
 NUMBER OF SEQUENCES: 20
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Cooper & Dunham LLP
 STREET: 1185 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 10036
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION NUMBER: US/08/757,669A
 APPLICATION DATA:
 FILING DATE:
 CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:
 NAME: White, John P.
 REGISTRATION NUMBER: 28,678
 REFERENCE/DOCKET NUMBER: 45185-F
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 278-0400
 TELEFAX: (212) 391-0525
 INFORMATION FOR SEQ ID NO: 20:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 32207 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 OS-08-757-669A-20

Query Match	6.73;	Score	121.2;	DB	4;	Length	32207;
Best Local Similarity	45.98;	Pred. No.	2.3e-18;				
Matches 452; Conservative	0;	Mismatches	573;	Indels	7;	Gaps	2

QY	299	MAGGAAGAACACACGCGGACACAGAAAGGGTTAAGTGGCTCACCGAGGACACAGTTG	352
DB	21015	ATTMGGATGATGACGAGAGACGACGACACAGTGTACGAGACGACGAGAGATATACGACGAG	20956
QY	353	GAGGACACAGTAAATCAGCACATATGTTAGAAATAGAGAAACGAGAGCGAAGACGAA	412
DB	20955	GATGACGAGACGAGATATACGAGGAGTTCGCGAGAGGATGACGAGAGGATATACGACGAGG	20896
QY	413	MAGGAATAGCATTCGTTATAGTGATGCGAGAAAGAACAGAGAGGAGGAACACAAAA	472

[illegible]

RESULT 6

US-09-230-371A-20/C
: Sequence 20, Application US/09230371A
: Patent No. 6348586
: GENERAL INFORMATION:
: APPLICANT: Chang, Yuan
: APPLICANT: Bohenzky, Roy A
: APPLICANT: Russo, James J
: APPLICANT: Edelman, Isidore S

QY 1450 CGACGAGAAAGCTCTATTCAGAGTAGTCGATCAGCGCAAAAGATCTTCAAGAACT 1507
 Db 2377 AAGTAGAAGAAAGAAAGAGAGAGAGACAGTAGAGAGAAAGATGTATTACAATT 2434

RESULT 10

US-08-719-822B-1
 ; Sequence 1, Application US/08719822B
 ; Patent No. 5874527

; GENERAL INFORMATION:
 ; APPLICANT: Barnwell, John
 ; TITLE OF INVENTION: Plasmodium vivax Blood Stage Antigens
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Darby and Darby
 ; STREET: 805 Third Ave.
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10022-7513

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/719,822B
 FILING DATE: 09/30/96
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Gogoris, Adda
 REGISTRATION NUMBER: 29,714
 REFERENCE/DOCKET NUMBER: 5986/17686US2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212)527-7700
 TELEFAX: (212)753-6237
 TELEX: 236687

INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3337 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Plasmodium vivax
 IMMEDIATE SOURCE:
 CLONE: PYMB3.3.1
 US-08-719-822B-1

Query Match 5.4%; Score 97.2; DB 2; Length 3337;
 Best Local Similarity 51.7%; Pred. No. 2,9e-13;
 Matches 247; Conservative 0; Mismatches 228; Indels 3; Gaps 1;

QY 1033 GACAGTACATCCAGAGAGCAAGCAAAACCAAGTAGAGAGAGTAGAGTGCAGC 1092
 Db 1957 GAGAGCCAAAG 2016
 QY 1093 TACAATAAACAAG 1152
 Db 2017 TCGAATTAGAAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2076
 QY 1153 TTCACAGTAG 1212
 Db 2077 CAG 2136
 QY 1213 TTCACAGTAG 1269
 Db 2137 CAG 2196
 QY 1270 TAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAG 1329

Db 2197 AAGAAAGTGAAG 2256
 QY 1330 GCACGAGAAAG 1389
 Db 2257 CAG 2316
 QY 1390 ATCTTCAAAAG 1449
 Db 2317 AGGTACCAAG 2376
 QY 1450 CGACGAGAAAGCTCTATTCAGAGTAGTCGATCAGCGCAAAAGATCTTCAAGAACT 1507
 Db 2377 AAGTAGAAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2434

RESULT 11

US-09-092-458-1
 ; Sequence 1, Application US/09092458
 ; Patent No. 6231861

; GENERAL INFORMATION:
 ; APPLICANT: Barnwell, John
 ; TITLE OF INVENTION: Plasmodium vivax Blood Stage Antigens, Monoclonal Antibodies, and Diagnostic Assays
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Darby and Darby
 ; STREET: 805 Third Ave.
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10022-7513

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/092,458
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/719,821
 FILING DATE: 09/30/96
 ATTORNEY/AGENT INFORMATION:
 NAME: Gogoris, Adda
 REGISTRATION NUMBER: 29,714
 REFERENCE/DOCKET NUMBER: 5986/17686US3
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212)527-7700
 TELEFAX: (212)753-6237
 TELEX: 236687

INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3337 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Plasmodium vivax
 IMMEDIATE SOURCE:
 CLONE: PYMB3.3.1
 US-09-092-458-1

Query Match 5.4%; Score 97.2; DB 4; Length 3337;
 Best Local Similarity 51.7%; Pred. No. 2,9e-13;
 Matches 247; Conservative 0; Mismatches 228; Indels 3; Gaps 1;

QY 1033 GACAGTACATCCAGAGAGCAAGCAAAACCAAGTAGAGAGAGTAGAGTGCAGC 1092
 Db 1957 GAGAGCCAAAG 2016
 QY 1093 TACAATAAACAAG 1152
 Db 2017 TCGAATTAGAAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2076
 QY 1153 TTCACAGTAG 1212
 Db 2077 CAG 2136
 QY 1213 TTCACAGTAG 1269
 Db 2137 CAG 2196
 QY 1270 TAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAG 1329


```

OY 1106 AGCAAGTAGTAGAGTGCAGAGCAGTAGCAGTCTAGTAGCAGTTCACACAGTACG 1165
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 61
OY 1166 AGCAAGTAGTAGTTCAGCAGAGTGCAGAGTACAGTCCAGTAGTTCAGTAGCAGC 1225
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 62 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 121
OY 1226 TCCAGTAGTAGTGCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1285
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 122 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 181
OY 1286 AGTAGAAGTCGAGTAG 1302
    ||| ||| ||| ||| |||
Db 182 AGCAGCAGCAGCAGCAGCAG 198

```

RESULT 14

```

5273901-6
: Patent No. 5273901
: APPLICANT: JACOBSON, JAMES W.; STRAUSBERG, ROBERT L.; WILSON,
: JESAM D.; POPE, SARON H.; STRAUSBERG, SUSAN L.; RUFF, MICHAEL D.;
: AUGUSTINE, PATRICIA C.; DANFORTH, HARRY D.
: TITLE OF INVENTION: GENETICALLY ENGINEERED COCCIDIOSIS
: SPOROZOITE 21.5 KB ANTIGEN, AC-6B
: NUMBER OF SEQUENCES: 11
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/581,693
: FILING DATE: 12-SEP-1990
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 215,162
: FILING DATE: 05-JUL-1988
: APPLICATION NUMBER: 746,520
: FILING DATE: 19-JUN-1985
: APPLICATION NUMBER: 627,811
: FILING DATE: 05-JUL-1984
: SEQ ID NO:6:
: LENGTH: 543
: 5273901-6

```

Query Match 5.0%; Score 89.8; DB 6; Length 543;

Best Local Similarity 56.7%; Pred. No. 7.1e-12; Matches 166; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

```

OY 1049 AGCAAGCAAAACCAAACTAGAGCAGAGAGTAGAGTGCAGCTAGCAATATAACAAGC 1108
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21 AGAAGAAAGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 80
OY 1109 AAGAGTAGAGTGCAGCAGTACAGTACAGTCTAGTAGCAGTTCACACAGTAGCAGC 1168
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    81 AAAAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 140
OY 1169 AGTGAAGTAGTTCAGCAGAGTGAAGTAGTAGCAGTGCAGTAGTTCAGTAGAGAGTCC 1228
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 141 AAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 200
OY 1229 AGTAGAAGTGCAGCAGCAGCAGAGATAGTAGCAGTAGCAGTACAGTAGTAGTAGAGT 1288
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 201 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 260
OY 1289 AGAAGTCGAGTAGGGGCTCGGGGACATATAGATAGAAAGCAGAGAGAGG 1341
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 261 ACCAGCTGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 313

```

RESULT 15

```

US-09-043-303-7
: Sequence 7, Application US/09043303
: Patent No. 6251589
: GENERAL INFORMATION:
: APPLICANT: TSOJI, Shoji
: APPLICANT: SANBEI, Kazuhiro
: TITLE OF INVENTION: Method for diagnosing Splincerebellar Ataxia Type 2 and

```

```

: TITLE OF INVENTION: Primers Therefor
: FILE REFERENCE: 0760-0241P
: CURRENT APPLICATION NUMBER: US/09/043,303
: CURRENT FILING DATE: 1998-05-18
: EARLIER APPLICATION NUMBER: PCT/JP96/01999
: EARLIER FILING DATE: 1996-07-18
: NUMBER OF SEQ ID NOS: 17
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 7
: LENGTH: 203
: TYPE: DNA
: ORGANISM: P-2093 plasmid
US-09-043-303-7

```

Query Match 4.8%; Score 86.6; DB 4; Length 203;

Best Local Similarity 65.0%; Pred. No. 2.6e-11; Matches 128; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

```

OY 1103 ACAAGCAGAGTAGAGTGCAGAGCAGCAGTACAGTACAGTCTAGTAGCAGTTCACACAGT 1162
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 64
OY 1163 AGCAGCAGTGCAGAGTTCAGCAGAGTGCAGAGTACAGTCCAGTAGTTCAGTAGC 1222
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 65 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 124
OY 1223 AGCTCAGTAGAGTGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1282
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 184
OY 1283 GAGAGTAGAAGTCGAG 1299
    ||| ||| ||| ||| |||
Db 185 ATCAGGAAACTCTGCG 201

```

Search completed: October 16, 2002, 13:06:56
Job time : 158 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 16, 2002, 12:49:43 ; Search time 310 Seconds
(without alignments)
9969.181 Million cell updates/sec

Title: US-09-822-849A-1

Perfect score: 1800
Sequence: 1 acgacagacacgacgacgtt.....ataaaaaagattacatcc 1800

Scoring table: IDENTITY_NUC
Gapop 10.0 ; Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Minimum number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: N_Geneseq_032802:.*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1730.4	96.1	2369	22	AAAF24910
2	607.4	33.7	616	21	AAA43765
3	393	21.8	1113	21	AAAF15981
4	299.4	16.6	325	20	AAV89627
5	268.4	14.9	293	16	AAAT20032
6	262.4	14.6	264	21	AAC02581
7	251.4	14.0	263	20	AAV86711
8	166.4	9.2	239	22	AAAS24061
9	166.4	9.2	239	22	AAH82631

10	136.4	7.6	3399	17	AAAT05868	Chicken leucocytos
11	121.2	6.7	3489	21	AAA30280	Kaposi's sarcoma-a
12	121.2	6.7	3489	22	AAAF82901	Nucleotide sequenc
13	121.2	6.7	3207	20	AAV73805	KSHV LTR DNA (nucl
14	121.2	6.7	137507	19	AAV19941	KSHV long unique c
15	118	6.6	1037	21	AAAS9242	Exons E, C and A o
16	118	6.6	1159	21	AAAS9240	An EcoRI fragment
17	118	6.6	1472	21	AAAS9241	Exons D, C, B and
18	116	6.4	1666	16	AAO87587	DNA encoding Leuco
19	103.6	5.8	5120	22	AAAC8467	DNA sequence of hu
20	102.6	5.7	3197	20	AAAC89891	Partial mouse WRN
21	99.2	5.5	16442	18	AAAX83006	P. vivax ESP-1 blo
22	97.2	5.4	3337	17	AAAT34620	DNA encoding a sec
23	97.2	5.4	3337	20	AAAT34620	Plasmid from vlvax E
24	97.2	5.4	3337	22	AAAT34620	Mouse ischaemic co
25	92.4	5.1	1080	24	AB199537	Human breast cell
26	91.6	5.1	575	22	ABA50472	Human foetal liver
27	91.6	5.1	575	22	ABA50472	Human foetal liver
28	91.6	5.1	575	22	ABA50472	Human foetal liver
29	91.6	5.1	575	22	ABA50472	Human foetal liver
30	91.6	5.1	575	22	ABA50472	Human foetal liver
31	91.6	5.1	575	22	ABA50472	Human foetal liver
32	91.6	5.1	575	22	ABA50472	Human foetal liver
33	91.6	5.1	575	22	ABA50472	Human foetal liver
34	91.6	5.1	1969	22	ABA55341	Human breast cell
35	91.6	5.1	1969	22	ABA55341	Human breast cell
36	91.6	5.1	1969	22	ABA55341	Human breast cell
37	91.6	5.1	1969	22	ABA55341	Human breast cell
38	91.6	5.1	1969	22	ABA55341	Human breast cell
39	91.6	5.1	1969	22	ABA55341	Human breast cell
40	91.6	5.1	1969	22	ABA55341	Human breast cell
41	91.6	5.1	1969	22	ABA55341	Human breast cell
42	90.8	5.0	8201	21	AAAB8864	Human dentin slalo
43	90.4	5.0	621	23	AAAS90688	DNA encoding novel
44	90.4	5.0	49999	20	AAZ23895	Murine LOBO homo
45	89.8	5.0	234	16	AAO84832	Spinocerebellar at

ALIGNMENTS

RESULT 1	
AAAF24910	
ID	AAAF24910 standard; cDNA: 2369 BP.
XX	AAAF24910;
XX	
DT	20-APR-2001 (first entry)
XX	
DE	Nucleotide sequence of a human pinin splice variant (PNIN).
XX	
KW	Human: Pinin splice variant; PNIN; vesicle trafficking; AIDS; emphysema;
KW	Addison's disease; neoplastic disorder; immunological disorder; gout;
KW	adult respiratory distress syndrome; allergy; anaemia; asthma; cancer;
KW	atherosclerosis; bronchitis; cholelithiasis; Crohn's disease; scleroderma;
KW	ulcerative colitis; atopic dermatitis; dermatomyositis; osteoporosis;
KW	glomerulonephritis; diabetes mellitus; erythema nodosum; pancreatitis;
KW	gastroenteritis; Graves' disease; hypereosinophilia; polymyositis;
KW	irritable bowel syndrome; lupus erythematosus; multiple sclerosis;
KW	myasthenia gravis; myocardial or pericardial inflammation; hemodialysis;
KW	osteoarthritis; rheumatoid arthritis; Sjogren's syndrome; trauma;
KW	Werner syndrome; autoimmune thyroiditis; infection; ss.
OS	Homo sapiens.
XX	
XX	
FT	Key
FT	CDS
FT	Location/Qualifiers
FT	25..2178
FT	/*tag- a
FT	/product= " pinin splice variant (PNIN) "
XX	
XX	US6162601-A.
XX	
PD	19-DEC-2000.

DB 2190 AGGCTTCTTACCATTTCTTGGACACAGACATTTCTGATTAAGAAAGCATTTCC 2249
 OY 1621 TTGTAAAGAGATGCTGCTTAAAGATTCATGTTGTAAGAAATCTTTTGGAAATACA 1680
 DB 2250 TTGTAAAGAGATGCTGCTTAAAGATTCATGTTGTAAGAAATCTTTTGGAAATACA 2309
 OY 1661 GACTGTGTTTACCAACATTTCTGTAATTTTGGATTAATTTTGAAGATTTATTC 1740
 DB 2310 GACTGTGTTTACCAACATTTCTGTAATTTTGGATTAATTTTGAAGATTTATTC 2369

RESULT 2

AAA43765
 ID AAA43765 standard; cDNA: 616 BP.

AC AAA43765;

DT 21-AUG-2000 (first entry)

XX Human secreted expressed sequence tag SEQ ID NO:340.

Human, mouse; chicken; rat; secreted expressed sequence tag; SEST;
 expressed sequence tag; EST; probe: chemokine; proliferative;
 immunomodulatory; haematopoietic; chemokine; analgesic; haemostatic;
 thrombolytic; antiinflammatory; cytosolic; antibacterial; antiinflammatory;
 antiviral; antidiabetic; antiproliferative; vulnery; antiparkinsonian;
 antitumor; osteoprotective; neuroprotective; nootropic; antiparasitic;
 cerebroprotective; anticonvulsant; antidepressant; gene therapy;
 vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
 insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
 lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
 central nervous system disorder; Alzheimer's disease; stroke;
 Parkinson's disease; Huntington's disease; coagulation disorder;
 haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
 tumor; infection; depression; psoriasis; ss.

XX Homo sapiens.

XX WO200021991-A1.

XX 20-APR-2000.

XX 15-OCT-1999; 99WO-US24206.

XX 15-OCT-1998; 98US-0104436.

XX (GENT) GENETICS INST INC.

XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;

XX Merberg D, Treacy M, Bowman MR;

XX WPI; 2000-317938/27.

XX Isolated polynucleotides, and encoded proteins, comprising secreted
 PT expressed sequence tags (SESTs), useful for treating various disorders
 PT such as autoimmune, infectious, and central nervous system disorders -
 PS claim 1; page 292; 803pp; English.

XX *AAA43426 to AAA45925 represent specifically claimed secreted expressed
 CC sequence tags (SESTs). Isolated from human, mouse, chicken and rat
 CC tissue sources. The SESTs can have a range of activities depending on
 CC the tissues they were isolated from. The activities include:
 CC chemokine; proliferative; immunomodulatory; haematopoietic;
 CC chemokine; analgesic; haemostatic; thrombolytic; antiinflammatory;
 CC cytosolic; antibacterial; antiinflammatory; antiviral; antidiabetic;
 CC antiasthmatic; vulnery; antiparkinsonian; osteoprotective;
 CC nootropic; antiparkinsonian; antiparasitic; cerebroprotective;
 CC anticonvulsant; and antidepressant. The SESTs can be used for gene
 CC therapy and in vaccines. The SESTs are useful as probes for the
 CC identification and isolation of full-length cDNAs and genomic DNA
 CC molecules which correspond to the SESTs. Proteins encoded by the SESTs

CC are useful in assays for determining biological activity and raising
 CC antibodies. They may be useful for treatment of autoimmune disorders
 CC (multiple sclerosis, insulin dependent diabetes), allergic conditions
 CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
 CC osteoporosis, osteoarthritis, central nervous system disorders
 CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
 CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
 CC disease), tumours, bacterial, fungal or viral infections, depression and
 CC psoriasis. AAA45926 to AAA45931 represent linker variants which are given
 CC in the exemplification of the present invention.

XX Sequence 616 BP: 211 A; 126 C; 154 G; 125 T; 0 other:

Query Match 33.7%; Score 607.4; DB 21; Length 616;

Best Local Similarity .99.8%; Pred. No. 1e-124;

Matches 608; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 467 CAAAAACAGAAATGAGATTAAGTATGAGAGAGAAATGAGTAAAGGAAAGTGAAG 526
 DB 2 CAAAAACAGAAATGAGATTAAGTATGAGAGAGAAATGAGTAAAGGAAAGTGAAG 61
 OY 527 CAGCAGATGATGACAGCTGAGAAAGTATGATGATGATGATGATGATGATGATGAT 586
 DB 62 CAGCAGATGATGACAGCTGAGAAAGTATGATGATGATGATGATGATGATGATGAT 121
 OY 587 CATGTAATGCTGACAGAGAGATTAATGAAATTAATGAAATGAAATGAAATGAAAT 646
 DB 122 CATGTAATGCTGACAGAGAGATTAATGAAATTAATGAAATGAAATGAAATGAAAT 181
 OY 647 GAAATGAAAGTATGAGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAAT 706
 DB 182 GAAATGAAAGTATGAGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAAT 241
 OY 707 GAATGAAATGCTTCTTCTGGAAGAGATGATGATGATGATGATGATGATGATGAT 766
 DB 242 GAATGAAATGCTTCTTCTGGAAGAGATGATGATGATGATGATGATGATGATGAT 301
 OY 767 TCTGAGAGAAAGAGAAAGAAATCTGAGCCCACTGAGCCCTGAGCTCAACTCAG 826
 DB 302 TCTGAGAGAAAGAGAAAGAAATCTGAGCCCACTGAGCCCTGAGCTCAACTCAG 361
 OY 827 CCTGAGTCTGAGCCCACTGAGCTTCAATCCAGTCCCAACAGTCTCAAGTCCAG 886
 DB 362 CCTGAGTCTGAGCCCACTGAGCTTCAATCCAGTCCCAACAGTCTCAAGTCCAG 421
 OY 887 CCTGAGTCTGAGCCCACTGAGCTTCAATCCAGTCTTCAATCCAGTCTTCAATCCAG 946
 DB 422 CCTGAGTCTGAGCCCACTGAGCTTCAATCCAGTCTTCAATCCAGTCTTCAATCCAG 481
 OY 947 CAGGAGCAAGGAGCATTTACTACCTGAGAGAGAGATTTCTGTAAGTCTGTAAGTCT 1006
 DB 482 CAGGAGCAAGGAGCATTTACTACCTGAGAGAGAGATTTCTGTAAGTCTGTAAGTCT 541
 OY 1007 ACTGAGTACCAAGTACAGCAGTCTGACAGTACCAAGTACCAAGTACCAAGTACCA 1066
 DB 542 ACTGAGTACCAAGTACAGCAGTCTGACAGTACCAAGTACCAAGTACCAAGTACCA 601
 OY 1067 ACTAGAGC 1075
 DB 602 ACCAGAGC 610

RESULT 3

AAFI5981
 ID AAFI5981 standard; cDNA: 1113 BP.

XX AAFI5981;

XX 13-MAR-2001 (first entry)

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:416.

XX Human; prostate cancer; prostate cancer antigen; detection; diagnosis;

neuroprotective; cytosolic; cardioactive; immunomodulatory; muscular;
 vulnerability; gastrointestinal; nephrotoxic; antineoplastic; gynaecological;
 antibacterial; gene therapy; neural; immune; reproductive; renal;
 gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
 wound; infectious disease; ss.

Homo sapiens.

MO200055174-A1.

21-SEP-2000.

08-MAR-2000; 2000MO-US05988.

12-MAR-1999; 99US-0124270.

(HOMA-) HUMAN GENOME SCI INC.

(ROSE/) ROSEN C A.

Rosen CA, Ruben SM;

MP1: 2000-587513/55.

P-PSDB: AAB56778.

Prostate cancer associated gene sequences, referred to as prostate cancer antigens, useful for treatment, prevention, and diagnosis of disorders such as prostate cancer -

Claim 1: Page 900-901; 238bp; English.

AA15566 to AA15605 encode the human prostate cancer associated proteins, called prostate cancer antigens, given in AAB56363 to AAB57302. The prostate cancer antigens can have neuroprotective, cytosolic, cardioactive, immunomodulatory, muscular, vulnerability, gastrointestinal, nephrotoxic, antineoplastic, gynaecological and antibacterial activities, and can be used in gene therapy. The prostate cancer antigen polynucleotides may be used for detection of prostate cancer, chromosome identification, as chromosome markers, and for numerous other diagnostic or research purposes. The prostate cancer antigens may be used to treat disorders such as neural, immune, muscular, reproductive, gastrointestinal, pulmonary, cardiovascular, renal, and proliferative disorders, wounds, and infectious diseases. AA15506 to AA15514 to CC AAB57303 represent sequences used in the exemplification of the present invention.

Sequence 1113 BP; 414 A; 189 C; 319 G; 190 T; 1 other;

Query Match 21.8%; Score 393; DB 21; Length 1113;

Best Local Similarity 96.4%; Pred. No. 3e-77; Mismatches 15; Indels 0; Gaps 0;

Matches 402; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

DB 1 ACAGACGAACTGGCGCTTTGGAAAGAAAGTTGAGCTGGCCAGCTGCAAGAAATG 60
 DB 658 ACAGACGAACTGGCGCTTTGGAAAGAAAGTTGAGCTGGCCAGCTGCAAGAAATG 717
 QY 61 GAATGAACATATGCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 120
 DB 718 GAATGAACATATGCAAAATTAATTAATTAATTAATTAATTAATTAATTAAT 777
 QY 121 TTTATTCCTGGAAGATGTCAGTACCAAAATTAATTAATTAATTAATTAATTAAT 180
 DB 778 TTTATTCCTGGAAGATGTCAGTACCAAAATTAATTAATTAATTAATTAATTAAT 837
 QY 181 AATGAACGCTTTATTTGAAGTAGACCATGCAATTTGCAAGCAAAATTAATTAATGA 240
 DB 838 AATGAACGCTTTATTTGAAGTAGACCATGCAATTTGCAAGCAAAATTAATTAATGA 897
 QY 241 GGTAGAGCTTAGAAGCAATCAATGAAGAAAAAGACATCAGTGGTGGCTAATGAGA 300
 DB 898 GGTAGAGCTTAGAAGCAATCAATGAAGAAAAAGACATCAGTGGTGGCTAATGAGA 957
 QY 301 ACAGAGCGGCAAGCAAGAGGTAAGTGGCTCAGCGAGAGAAAGTTGAGAGAGAC 360

DB 958 ACAGAGCGGCAAGCAAGAGGTAAGTGGCTCAGCGAGAGAAAGTTGAGAGAGAC 1017
 QY 361 AGGTAAATACACATATGATGTAAGAAATAGAGAAAGCAAGAGAGAAAGAAAGCA 417
 DB 1018 AGGTAAATACACATATGATGTAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 1074

RESULT 4

AAV89627 standard; cDNA; 325 BP.

AAV89627;

15-FEB-1999 (first entry)

EST clone COS59.

Human; secreted protein; expressed sequence tag; EST; haematopoiesis; tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic; receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour; gene therapy; ss.

Homo sapiens.

MO9845436-A2.

15-OCT-1998.

10-APR-1998; 98MO-US06955.

10-APR-1997; 97US-0838821.

(GENY) GENETICS INST INC.

Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;

Racie LA, Spaulding V, Treacy M;

MP1: 1999-070077/06.

New polynucleotides encoding human secreted proteins - derived from e.g. human blood, kidney, foetal lung, placenta, testes, brain, ovary, pituitary, retina and colon cDNA libraries.

Claim 1: Page 278; 618bp; English.

The present sequence represents a human expressed sequence tag (EST). The polynucleotide, which is a secreted EST, and the encoded protein are predicted to have useful biological activities which would make them suitable for treating, preventing or ameliorating medical conditions in humans and animals, although no supporting data is given. Suggested activities include nutritional activity, immune stimulating or suppressing activity, haematopoiesis regulating activity, tissue growth activity, activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, receptor/ligand activity, anti-inflammatory activity, cadherin/tumour invasion suppressor activity, tumour inhibition activity. The polynucleotide may also be useful for gene therapy.

Sequence 325 BP; 123 A; 47 C; 87 G; 68 T; 0 other;

Query Match 16.6%; Score 299.4; DB 20; Length 325;

Best Local Similarity 99.7%; Pred. No. 9.3e-57; Mismatches 1; Indels 0; Gaps 0;

Matches 300; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 504 CTGAGTAAGGCAAGTGAAGACAGCAGATAGTACGCTTGAAGAGTTATGATGTC 563
 DB 18 CTGAGTAAGGCAAGTGAAGACAGCAGATAGTACGCTTGAAGAGTTATGATGTC 77
 QY 564 TAGAGATGCTTGAAGATGCAACATGTAATGCTACAGAGGAGTAATGAACTAATC 623
 DB 78 TAGAGATGCTTGAAGATGCAACATGTAATGCTACAGAGGAGTAATGAACTAATC 137
 QY 624 GAGTTAAAGGTAGAACCTTCAGAAAATGAAGCTAGCAAGAAATGGAACCAAGATG 683

```

|||||
Db 138 GAGTTGAAAGTGTAGACCTTCAGAAATGAAGCTAGCAAAAGATTGAAACCAAGAAATG 197
Oy 684 ATTGGAATTTAGCCAGATTAAGATGTAATCCCTTCTCTCGGGAAGAGATGCA 743
Db 198 ATTGGAATTTAGCCAGATTAAGATGTAATCCCTTCTCTCGGGAAGAGATGCA 257
Oy 744 GTGCTTTAGACATGGAAGAGCTGAGGAGAGAGAGAGAGAGAGAGATGAGCCCAAC 803
Db 258 GTGCTTTAGACATGGAAGAGCTGAGGAGAGAGAGAGAGAGAGATGAGCCCAAC 317
Oy 804 C 804
Db 318 C 318

```

RESULT 5

```

AAT20032
ID AAT20032 standard: cDNA to mRNA: 293 BP.
Oy AAT20032:
Db 17-JUL-1996 (first entry)

```

Human gene signature HUMGS01171.

Gene signature: messenger RNA; mRNA; relative abundance; frequency; human; cloning; mapping; non-biased library; diagnosis; detection; cell typing; abnormal cell function; ss.

Homo sapiens.

MO9514772-AL.

01-JUN-1995.

11-NOV-1994: 94MO-JP01916.

12-NOV-1993: 93JP-0355504.

(MATS/) MATSUBARA K.

(OKUB/) OKUBO K.

Matsubara K, Okubo K;

WPI: 1995-206931/27.

Identifying gene signatures in 3'-directed human cDNA library - e.g. for diagnosis of abnormal cell function, by preparing cDNA that reflects relative abundance of corresp. mRNA in specific human tissues

Claim 1: Page 542: 2245pp; Japanese.

A single-stranded DNA (or its complementary strand or the corresp. double-stranded DNA) which comprises one of the 7837 "GS" sequences given in AAT19001-T26837 and which is able to hybridise to part of human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature) sequences were obtained from 3'-directed cDNA libraries prepared from various human tissues; synthesis of cDNA was initiated from the 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-untranslated sequence is unique to a particular mRNA species, almost all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library is constructed so as to reflect accurately the relative abundance of different mRNAs in the particular tissue from which it was derived. The appearance frequency of a given GS in a cDNA library can be determined (esp. using primers and probes derived from the GS sequences) as a means of diagnosing abnormal cell function or for recognising different cell types.

Sequence 293 BP: 103 A; 38 C; 57 G; 80 T; 15 other;

Query Match

14.9%; Score 268.4; DB 16; Length 293;

Best Local Similarity 93.2%; Pred. No. 6, 3e-50; Matches 272; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

```

Oy 1494 GATCTTCAGAGAGTGAAGAGCCGAGAAATCAGACAGAGAGAGAGAGAGAGAGTGA 1553
Db 1 GATCTTCAGAGAGTGAAGAGCCGAGAAATCAGACAGAGAGAGAGAGAGAGTGA 60
Oy 1554 AGAAGCCAGGCTTTCTTACCATTTCTTGACAGAGAGATTTCTTGATTAAGAGATTA 1613
Db 61 AGAAGCCAGGCTTTCTTACCATTTCTTGACAGAGAGATTTCTTGATTAAGAGATTA 120
Oy 1614 CCTTCCTTTAAGAGAGATGCTGCTTAAGATTCGATGTAAGAGATTTCTTGCA 1673
Db 121 CCTTCCTTTAAGAGAGATGCTGCTTAAGATTCGATGTAAGAGATTTCTTGCA 180
Oy 1674 AATACAGACTGTTGTTTACAGACATTTCTTACTTTTGCATATTTTGAAGATT 1733
Db 181 AATACAGACTGTTGTTTACAGACATTTCTTACTTTTGCATATTTTGAAGATT 240
Oy 1734 ATTTATCAAAATTTATGTGAGCTTCCAAAATATGTAAATGATATATATAA 1785
Db 241 ATTTATCAAAATTTATGTGAGCTTCCAAAATATGTAAATGATATATATAA 292

```

RESULT 6

AAC02581

AAC02581:

06-OCT-2000 (first entry)

Human secreted protein 5' EST, SEQ ID NO: 2579.

Human: 5' EST; expressed sequence tag; secreted protein; cDNA isolation; gene therapy; chromosome mapping; ss.

Homo sapiens.

EP1033401-A2.

06-SEP-2000.

21-FEB-2000; 2000EP-0200610.

26-FEB-1999; 99US-0122487.

(GEST) GENSET.

Dumas Mline Edwards J, Duclert A, Giordano J;

WPI: 2000-500381/45.

P-PSDB: AAG02575.

New nucleic acid that is a 5' expressed sequence tag (5' EST) for obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for diagnostic, forensic, gene therapy and chromosome mapping procedures - Claim 1: SEQ ID 2579; 71pp + CD-ROM; English.

The present sequence is one of a large number of 5' ESTs derived from mRNAs encoding secreted proteins. An ORF has been identified within the sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs derived from 30 different tissues. EST sequences usually correspond mainly to the 3' untranslated region (UTR) of the mRNA because they are often obtained from oligo-dT primed cDNA libraries. Such ESTs are not well suited for isolating cDNA sequences derived from the 5' ends of cDNAs and even in those cases where longer cDNA sequences have been obtained, the full 5' UTR is rarely included. 5' ESTs are derived from cDNAs with intact 5' ends and can therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and chromosome mapping procedures. They are used to obtain upstream regulatory sequences and to design expression and secretion

CC vectors.

Sequence 264 BP: 80 A; 73 C; 53 G; 58 T; 0 other;

Query Match 14.6%; Score 262.4; DB 21; Length 264;

Best Local Similarity 99.6%; Pred. No. 1.3e-48;

Matches 263; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

677 GAATGGAATTTAAATGAGCCAGATTAAGATTAATTCCTTCTCTGGAAAGAG 736

1 GAATGGAATTTAAATGAGCCAGATTAAGATTAATTCCTTCTCTGGAAAGAG 60

737 AATGTCAGTCTTTAGACATGAAAAGAGCTGAGAAAAAGAAAAGATCTGAG 736

61 AATGTCAGTCTTTAGACATGAAAAGAGCTGAGAAAAAGAAAAGATCTGAG 120

797 CCGCAAGCTAGGCTGCTGCTCAAGCTCAAGCTCAAGCTCAAGCTCAAGCTCA 856

121 CCGCAAGCTAGGCTGCTGCTCAAGCTCAAGCTCAAGCTCAAGCTCAAGCTCA 180

857 TCCAGTCCCAACAGTACTCCAGTCCCAAGTCCCAAGTCCCAAGTCCCAAGTCC 916

181 TCCAGTCCCAACAGTACTCCAGTCCCAAGTCCCAAGTCCCAAGTCCCAAGTCC 240

917 GCTGTTTAAAGCCACACCCCA 940

241 GCTGTTTAAAGCCACACCCCA 264

RESULT 7

AAV86711 standard; cDNA; 263 BP.

AAV86711:

27-APR-1999 (first entry)

EST clone AM32.

Expressed sequence tag; secreted protein; haematopoiesis regulator;

tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;

chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;

receptor; ligand; anti-inflammatory; tumour inhibitor; ds.

Homo sapiens.

MO9845435-A2.

15-OCT-1998.

10-APR-1998; 98MO-US06954.

10-APR-1997; 97US-0835913.

(GEMT) GENETICS INST INC.

Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Werberg D;

Racle LA, Spaulding V, Treacy M;

WPI; 1999-070076/06.

New polynucleotides encoding human secreted proteins - derived from

e.g. human blood, kidney, foetal lung, placenta, testes, brain,

ovary, pituitary, retina and colon cDNA libraries

Claim 1, Page 324, 633pp; English.

CC animals, although no supporting data is given. Suggested activities

include nutritional activity, immune stimulating or suppressing activity,

haematopoiesis regulating activity, tissue growth activity,

activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic

and thrombolytic activity, receptor/ligand activity, anti-inflammatory

activity, cadherin/tumour invasion suppressor activity, tumour inhibition

therapy. The EST sequences are also stated to be useful for gene

therapy.

Sequence 263 BP: 107 A; 38 C; 76 G; 42 T; 0 other;

Query Match 14.0%; Score 251.4; DB 20; Length 263;

Best Local Similarity 99.6%; Pred. No. 3.5e-46;

Matches 252; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

127 TCCGTGAAGATGTGTCAGCTCCCAAACTATATGAAAGTCAAGAAATGAA 186

2 TCCGTGAAGATGTGTCAGCTCCCAAACTATATGAAAGTCAAGAAATGAA 61

187 CCGTTTATTTGAAGTGAAGCATTCGAATTTGCGAAGCAAAATTAATGAGCTAG 246

62 CCGTTTATTTGAAGTGAAGCATTCGAATTTGCGAAGCAAAATTAATGAGCTAG 121

247 GCCTAGAACCAATCAATGAGCAAAAGAGCATTCAGTGTGCTGTAATGAAGACAGAA 306

122 GCCTAGAACCAATCAATGAGCAAAAGAGCATTCAGTGTGCTGTAATGAAGACAGAA 181

307 GCGGAAACAAGAGAGGTAGGTGCTGCGAGAGAGAAAGTTGAGAGACAGGTAA 366

182 GCGGAAACAAGAGAGGTAGGTGCTGCGAGAGAGAAAGTTGAGAGACAGGTAA 241

367 TCAAGCAATGAT 379

242 TCAAGCAATGAT 254

RESULT 8

AA524061 standard; cDNA; 239 BP.

AA524061:

07-NOV-2001 (first entry)

Human ovarian PCR-subtracted cDNA library clone #242.

Immunogenic protein; cancer; ovarian tumour; T-cell stimulation; ss;

gene therapy; cytostatic; T-cell expansion; nucleic acid hybridisation;

primer; probe.

Homo sapiens.

MO200157207-A2.

09-AUG-2001.

05-FEB-2001; 2001MO-US03733.

04-FEB-2000; 2000US-0180403.

28-MAR-2000; 2000US-0192745.

(CORI-) CORIXA CORP.

Algate PA, Mannion J;

WPI; 2001-488879/53.

New polynucleotides encoding ovarian tumour proteins, useful for

treating ovarian cancer, and as probes, primers, and markers of cancer

progression

Example 1; page 149; 378pp; English.

CC The invention comprises compositions used for the therapy and diagnosis
CC of ovarian cancer. The compositions comprise one or more ovarian tumour
CC proteins, their associated polynucleotides, or immunogenic portions of
CC the proteins. The ovarian tumour polynucleotides and polypeptides are
CC useful for stimulating and/or expanding T cells specific for a tumour
CC protein. They are also useful for inhibiting the development of cancer in
CC a patient with an ovarian tumour DNA or protein by incubating isolated
CC T-cells allowing them to proliferate, and administering to the patient.
CC The sequences can be used as markers for cancer, for example, to monitor
CC ovarian cancer progression. Probes and primers are useful in nucleic acid
CC hybridisation, in detecting the presence of complementary sequences in a
CC given sample, for preparing mutant species and for preparing other
CC genetic constructions. Sequences AA23820-AA25231 and AA25328-AA25549
CC represent human ovarian tumour protein cDNA clones.

XX Sequence 239 BP; 81 A; 28 C; 74 G; 42 T; 14 other;

Query Match 9.2%; Score 166.4; DB 22; Length 239;
Best Local Similarity 93.4%; Pred. No. 2.1e-27;
Matches 170; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Db 1203 GTCCAGTAGTTCAGTAGACGCTCCAGTACAGTGCAGCAGCAGAGATAGTACCA 1262
|||||
1 GTCCAGTAGTTCAGTAGACGCTCCAGTACAGTGCAGCAGCAGAGATAGTACCA 60

Qy 1263 GTACCACTAGTAGTAGTAGAGAGTAGAGGCGGCGGACATATATAGAG 1322
|||||
Db 61 GTACCACTAGTAGTAGTAGAGAGTAGAGGCGGCGGACATATATAGAG 120

Qy 1333 ATAGAAACACAGAAAGAGCGGTGATCGAAGAGAGGATATTCTCAGGACTAGAAAGAA 1382
|||||
Db 121 ATAGAAACACAGAAAGAGCGGTGATCGAAGAGGATATTCTCAGGAAATGGAAGAT 180

Qy 1383 GT 1384
||
Db 181 GT 182

RESULT 9
AAH82631
ID AAH82631 standard; cDNA: 239 BP.

XX AAH82631;
AC
XX 25-SEP-2001 (first entry)

DE Human ovarian tumour associated polynucleotide sequence SEQ ID NO:255.
XX Human; ovarian tumour; ovarian cancer; diagnosis; gene therapy;
XX immunogenic; vaccine; ss.

XX Homo sapiens.
OS
XX MO200151513-A2.

XX 19-JUL-2001.
PD
XX 16-JAN-2001; 2001MO-US01575.

XX 14-JAN-2000; 2000US-0176722.
PR
XX (CORI-) CORIXA CORP.

XX Algate PA;
PA
XX WPI: 2001-425866/45.

XX Novel ovarian tumor proteins, and nucleic acids encoding them, used to
XX treat and diagnose cancers, particularly ovarian cancer -
XX Claim 5; Page 109; 338pp; English.

XX AAH82377 to AAH83878 represent human ovarian tumour-associated

CC polynucleotide sequences which encode ovarian tumour proteins. The
CC ovarian tumour protein and polynucleotide sequences have cytostatic
CC activity, and can be used in gene therapy and vaccine production. The
CC ovarian tumour proteins and polynucleotides can be used to inhibit
CC the development of cancer, particularly ovarian cancer. They can also
CC be used to diagnose the onset and progression of cancer.

XX Sequence 239 BP; 81 A; 28 C; 74 G; 42 T; 14 other;

Query Match 9.2%; Score 166.4; DB 22; Length 239;
Best Local Similarity 93.4%; Pred. No. 2.1e-27;
Matches 170; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 1203 GTCCAGTAGTTCAGTAGACGCTCCAGTACAGTGCAGCAGCAGAGATAGTACCA 1262
|||||
Db 1 GTCCAGTAGTTCAGTAGACGCTCCAGTACAGTGCAGCAGCAGAGATAGTACCA 60

Qy 1263 GTACCACTAGTAGTAGTAGAGAGTAGAGGCGGCGGACATATATAGAG 1322
|||||
Db 61 GTACCACTAGTAGTAGTAGAGAGTAGAGGCGGCGGACATATATAGAG 120

Qy 1333 ATAGAAACACAGAAAGAGCGGTGATCGAAGAGAGGATATTCTCAGGACTAGAAAGAA 1382
|||||
Db 121 ATAGAAACACAGAAAGAGCGGTGATCGAAGAGGATATTCTCAGGAAATGGAAGAT 180

Qy 1383 GT 1384
||
Db 181 GT 182

RESULT 10
AAT05868
ID AAT05868 standard; DNA: 3399 BP.

XX AAT05868;
AC
XX 14-AUG-1996 (first entry)

DE Chicken leucocytozoan DNA encoding immunogenic protein for vaccines.
XX Chicken leucocytozoan; immunogen; recombinant vaccine; protection;
XX immunisation; vaccination; ss.

XX Chicken leucocytozoan.
OS
XX Key Location/Qualifiers
FT CDS 1..3399
FT /*tag- a
FT 1150..3218
FT /*tag- b
FT /note- "fragment referred to in the claims, for
FT use as insert in a recombinant vaccine
FT against chicken leucocytozoan disease"

XX JP07284392-A.
PN
XX 31-OCT-1995.
PD
XX 19-APR-1994; 94JP-0080643.

XX 19-APR-1994; 94JP-0080643.
PR
XX (DOBU-) DOBUTSUYO SEIBUTSUGAKUTERI SEIZAI KYOKAI.
XX (KITA) KITASATO KENKYUSHO SH.

XX WPI: 1996-006311/01.
XX P-PSDB: AAR97866.

XX Chicken leucocytozoan immunogenic protein - used in a recombinant
XX vaccine against chicken leucocytozoan disease
XX Claim 6; Page 6-9; 35pp; Japanese.

XX

